4.2.1 SOAR

The cognitive architectures best known among AI academics are probably Soar and ACT-R, both of which are explicitly being developed with the dual goals of creating human-level AGI and modeling all aspects of human psychology. Neither the Soar nor ACT-R communities feel themselves particularly near these long-term goals, yet they do take them seriously.

Soar is based on IF-THEN rules, otherwise known as "production rules." On the surface this makes it similar to old-style expert systems, but Soar is much more than an expert system; it's at minimum a sophisticated problem-solving engine. Soar explicitly conceives problem solving as a search through solution space for a "goal state" representing a (precise or approximate) problem solution. It uses a methodology of incremental search, where each step is supposed to move the system a little closer to its problem-solving goal, and each step involves a potentially complex "decision cycle."

In the simplest case, the decision cycle has two phases:

- Gathering appropriate information from the system's long-term memory (LTM) into its working memory (WM)
- A decision procedure that uses the gathered information to decide an action

If the knowledge available in LTM isn't enough to solve the problem, then the decision procedure invokes search heuristics like hill-climbing, which try to create new knowledge (new production rules) that will help move the system closer to a solution. If a solution is found by chaining together multiple production rules, then a chunking mechanism is used to combine these rules together into a single rule for future use. One could view the chunking mechanism as a way of converting explicit knowledge into implicit knowledge, similar to "map formation" in CogPrime (see Chapter 42 of Part 2), but in the current Soar design and implementation it is a fairly crude mechanism.

In recent years Soar has acquired a number of additional methods and modalities, including some visual reasoning methods and some mechanisms for handling episodic and procedural knowledge. These expand the scope of the system but the basic production rule and chunking mechanisms as briefly described above remain the core "cognitive algorithm" of the system.

From a CogPrime perspective, what Soar offers is certainly valuable, e.g.

- heuristics for transferring knowledge from LTM into WM
- chaining and chunking of implications
- methods for interfacing between other forms of knowledge and implications

However, a very short and very partial list of the major differences between Soar and CogPrime would include

- CogPrime contains a variety of other core cognitive mechanisms beyond the management and chunking of implications
- the variety of "chunking" type methods in CogPrime goes far beyond the sort of localized chunking done in Soar
- CogPrime is committed to representing uncertainty at the base level whereas Soar's production rules are crisp
- The mechanisms for LTM-WM interaction are rather different in CogPrime, being based on complex nonlinear dynamics as represented in Economic Attention Allocation (ECAN)
- Currently Soar does not contain creativity-focused heuristics like blending or evolutionary learning in its core cognitive dynamic.

4.2.2 ACT-R

In the grand scope of cognitive architectures, ACT-R is quite similar to Soar, but there are many micro-level differences. ACT-R is defined in terms of declarative and procedural knowledge, where procedural knowledge takes the form of Soar-like production rules, and declarative knowledge takes the form of chunks. It contains a variety of mechanisms for learning new rules and chunks from old; and also contains sophisticated probabilistic equations for updating the activation levels associated with items of knowledge (these equations being roughly analogous in function to, though quite different from, the ECAN equations in CogPrime).

Figure 4.2 displays the current architecture of ACT-R. The flow of cognition in the system is in response to the current goal, currently active information from declarative memory, information attended to in perceptual modules (vision and audition are implemented), and the current state of motor modules (hand and speech are implemented). The early work with ACT-R was based on comparing system performance to human behavior, using only behavioral measures, such as the timing of keystrokes or patterns of eye movements. Using such measures, it was not possible to test detailed assumptions about which modules were active in the performance of a task. More recently the ACT-R community has been engaged in a process of using imaging data to provide converging data on module activity. Figure 4.3 illustrates the associations they have made between the modules in Figure 4.2 and brain regions. Coordination among all of these components occurs through actions of the procedural module, which is mapped to the basal ganglia.

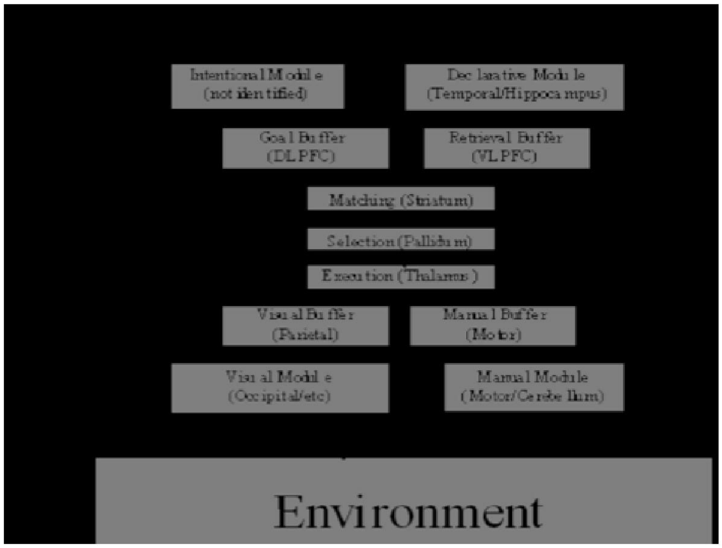

Intentional Module
(not identified)
Declarative Module
(Temporal/Hippocampus)
Goal Buffer
(DLPFC)
Retrieval Buffer
(VLPFC)
Matching (Striatum)
Selection (Pallidum)
Execution (Thalamus)
Visual Buffer
(Parietal)
Manual Buffer
(Motor)
Visual Module
(Occipital/etc)
Manual Module
(Motor/Cerebellum)
Environment

Fig. 4.2: High-level architecture of ACT-R

In practice ACT-R, even more so than Soar, seems to be used more as a programming framework for cognitive modeling than as an AI system. One can fairly easily use ACT-R to program models of specific human mental behaviors, which may then be matched against

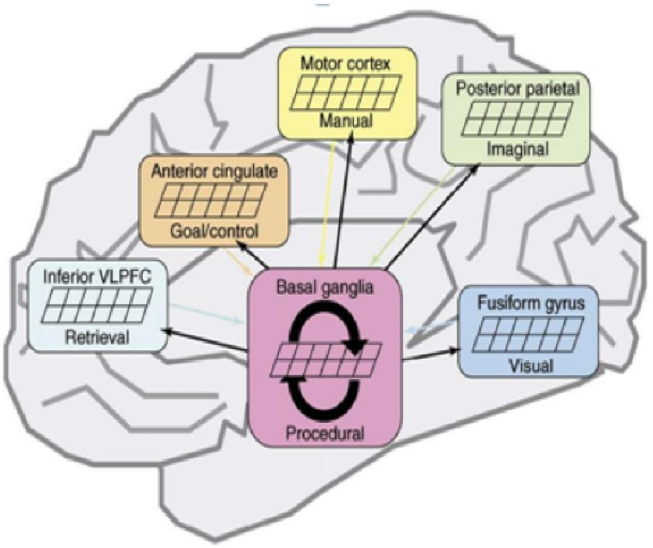


Fig. 4.3: Conjectured Mapping Between ACT-R and the Brain

psychological data. Opinions differ as to whether this sort of modeling is valuable for achieving AGI goals. CogPrime is not designed to support this kind of modeling, as it intentionally does many things very differently from humans.

ACT-R in its original form did not say much about perceptual and motor operations, but recent versions have incorporated EPIC, an independent cognitive architecture focused on modeling these aspects of human behavior.

4.2.3 Cyc and Texai

Our review of cognitive architectures would be incomplete without mentioning Cyc [LG90], one of the best known and best funded AGI-oriented projects in history. While the main focus of the Cyc project has been on the hand-coding of large amounts of declarative knowledge, there is also a cognitive architecture of sorts there. The center of Cyc is an engine for logical deduction, acting on knowledge represented in predicate logic. A natural language engine has been associated with the logic engine, which enables one to ask English questions and get English replies.

Stephen Reed, while an engineer at Cycorp, designed a perceptual-motor front end for Cyc based on James Albus' Reference Model Architecture; the ensuing system, called CognitiveCyc, would have been the first full-fledged cognitive architecture based on Cyc, but was not implemented. Reed left Cycorp and is now building a system called Texai, which has many similarities to Cyc (and relies upon the OpenCyc knowledge base, a subset of Cyc's overall knowledge base), but incorporates a CognitiveCyc style cognitive architecture.

4.2.4 NARS

Pei Wang's NARS logic [Wan06] played a large role in the development of PLN, CogPrime's uncertain logic component, a relationship that is discussed in depth in [GMIH08] and won't be re-emphasized here. However, NARS is more than just an uncertain logic, it is also an overall cognitive architecture (which is centered on NARS logic, but also includes other aspects). CogPrime bears little relation to NARS except in the specific similarities between PLN logic and NARS logic, but, the other aspects of NARS are worth briefly recounting here.

NARS is formulated as a system for processing tasks, where a task consists of a question or a piece of new knowledge. The architecture is focused on declarative knowledge, but some pieces of knowledge may be associated with executable procedures, which allows NARS to carry out control activities (in roughly the same way that a Prolog program can).

At any given time a NARS system contains

- working memory: a small set of tasks which are active, kept for a short time, and closely related to new questions and new knowledge
- long-term memory: a huge set of knowledge which is passive, kept for a long time, and not necessarily related to current questions and knowledge

The working and long term memory spaces of NARS may each be thought of as a set of chunks, where each chunk consists of a set of tasks and a set of knowledge. NARS's basic cognitive process is:

1. choose a chunk
2. choose a task from that chunk
3. choose a piece of knowledge from that chunk
4. use the task and knowledge to do inference
5. send the new tasks to corresponding chunks

Depending on the nature of the task and knowledge, the inference involved may be one of the following:

- if the task is a question, and the knowledge happens to be an answer to the question, a copy of the knowledge is generated as a new task
- backward inference
- revision (merging two pieces of knowledge with the same form but different truth value)
- forward inference
- execution of a procedure associated with a piece of knowledge

Unlike many other systems, NARS doesn't decide what type of inference is used to process a task when the task is accepted, but works in a data-driven way – that is, it is the task and knowledge that dynamically determine what type of inference will be carried out

The "choice" processes mentioned above are done via assigning relative priorities to

- chunks (where they are called activity)
- tasks (where they are called urgency)
- knowledge (where they are called importance)

and then distributing the system's resources accordingly, based on a probabilistic algorithm. (It's interesting to note that while NARS uses probability theory as part of its control mechanism, the logic it uses to represent its own knowledge about the world is nonprobabilistic. This is considered conceptually consistent, in the context of NARS theory, because system control is viewed as a domain where the system's knowledge is more complete, thus more amenable to probabilistic reasoning.)

4.2.5 GLAIR and SNePS

Another logic-focused cognitive architecture, very different from NARS in detail, is Stuart Shapiro's GLAIR cognitive architecture, which is centered on the SNePS paraconsistent logic [SE07].

Like NARS, the core "cognitive loop" of GLAIR is based on reasoning: either thinking about some percept (e.g. linguistic input, or sense data from the virtual or physical world), or answering some question. This inference based cognition process is turned into an intelligent agent control process via coupling it with an acting component, which operates according to a set of policies, each one of which tells the system when to take certain internal or external actions (including internal reasoning actions) in response to its observed internal and external situation.

GLAIR contains multiple layers:

- the Knowledge Layer (KL), which contains the beliefs of the agent, and is where reasoning, planning, and act selection are performed
- the Sensori-Actuator Layer (SAL), contains the controllers of the sensors and effectors of the hardware or software robot.
- the Perceptuo-Motor Layer (PML), which grounds the KL symbols in perceptual structures and subconscious actions, contains various registers for providing the agent's sense of situatedness in the environment, and handles translation and communication between the KL and the SAL.

The logical Knowledge Layer incorporates multiple memory types using a common representation (including declarative, procedural, episodic, attentional and intentional knowledge, and meta-knowledge). To support this broad range of knowledge types, a broad range of logical inference mechanisms are used, so that the KL may be variously viewed as predicate logic based, frame based, semantic network based, or from other perspectives.

What makes GLAIR more robust than most logic based AI approaches is the novel paraconsistent logical formalism used in the knowledge base, which means (among other things) that uncertain, speculative or erroneous knowledge may exist in the system's memory without leading the system to create a broadly erroneous view of the world or carry out egregiously unintelligent actions. CogPrime is not thoroughly logic-focused like GLAIR is, but in its logical aspect it seeks a similar robustness through its use of PLN logic, which embodies properties related to paraconsistency.

Compared to CogPrime, we see that GLAIR has a similarly integrative approach, but that the integration of different sorts of cognition is done more strictly within the framework of logical knowledge representation.

4.3 Emergentist Cognitive Architectures

Another species of cognitive architecture expects abstract symbolic processing to emerge from lower-level "subsymbolic" dynamics, which sometimes (but not always) are designed to simulate neural networks or other aspects of human brain function. These architectures are typically strong at recognizing patterns in high-dimensional data, reinforcement learning and associative memory; but no one has yet shown how to achieve high-level functions such as abstract reasoning or complex language processing using a purely subsymbolic approach. A few of the more important subsymbolic, emergentist cognitive architectures are:

- **DeSTIN** [ARK09a, ARC09], which is part of CogPrime, may also be considered as an autonomous AGI architecture, in which case it is emergentist and contains mechanisms to encourage language, high-level reasoning and other abstract aspects of intelligent to emerge from hierarchical pattern recognition and related self-organizing network dynamics. In CogPrime DeSTIN is used as part of a hybrid architecture, which greatly reduces the reliance on DeSTIN's emergent properties.
- **Hierarchical Temporal Memory (HTM)** [HB06] is a hierarchical temporal pattern recognition architecture, presented as both an AI approach and a model of the cortex. So far it has been used exclusively for vision processing and we will discuss its shortcomings later in the context of our treatment of DeSTIN.
- **SAL** [JL08], based on the earlier and related **IBCA** (Integrated Biologically-based Cognitive Architecture) is a large-scale emergent architecture that seeks to model distributed information processing in the brain, especially the posterior and frontal cortex and the hippocampus. So far the architectures in this lineage have been used to simulate various human psychological and psycholinguistic behaviors, but haven't been shown to give rise to higher-level behaviors like reasoning or subgoaling.
- **NOMAD** (Neurally Organized Mobile Adaptive Device) automata and its successors [KE06] are based on Edelman's "Neural Darwinism" model of the brain, and feature large numbers of simulated neurons evolving by natural selection into configurations that carry out sensorimotor and categorization tasks. The emergence of higher-level cognition from this approach seems rather unlikely.
- Ben Kuipers and his colleagues [MK07, MK08, MK09]have pursued an extremely innovative research program which combines qualitative reasoning and reinforcement learning to enable an intelligent agent to learn how to act, perceive and model the world. Kuipers' notion of "bootstrap learning" involves allowing the robot to learn almost *everything* about its world, including for instance the structure of 3D space and other things that humans and other animals obtain via their genetic endowments. Compared to Kuipers' approach, CogPrime falls in line with most other approaches which provide more "hard-wired" structure, following the analogy to biological organisms that are born with more innate biases.

There is also a set of emergentist architectures focused specifically on developmental robotics, which we will review below in a separate subsection, as all of these share certain common characteristics.

Our general perspective on the emergentist approach is that it is philosophically correct but currently pragmatically inadequate. Eventually, *some* emergentist approach could surely succeed at giving rise to humanlike general intelligence – the human brain, after all, is plainly an emergentist system. However, we currently lack understanding of how the brain gives rise to abstract reasoning and complex language, and none of the existing emergentist systems seem remotely capable of giving rise to such phenomena. It seems to us that the creation of a successful emergentist AGI will have to wait for either a detailed understanding of how the brain gives rise to abstract thought, or a much more thorough mathematical understanding of the dynamics of complex self-organizing systems.

The concept of cognitive synergy is more relevant to emergentist than to symbolic architectures. In a complex emergentist architecture with multiple specialized components, much of the emergence is expected to arise via synergy between different richly interacting components. Symbolic systems, at least in the forms currently seen in the literature, seem less likely to give rise to cognitive synergy as their dynamics tend to be simpler. And hybrid systems, as we shall see, are somewhat diverse in this regard: some rely heavily on cognitive synergies and others consist of more loosely coupled components.

We now review the DeSTIN emergentist architecture in more detail, and then turn to the developmental robotics architectures.

4.3.1 DeSTIN: A Deep Reinforcement Learning Approach to AGI

The DeSTIN architecture, created by Itamar Arel and his colleagues, addresses the problem of general intelligence using hierarchical spatiotemporal networks designed to enable scalable perception, state inference and reinforcement-learning-guided action in real-world environments. DeSTIN has been developed with the plan of gradually extending it into a complete system for humanoid robot control, founded on the same qualitative information-processing principles as the human brain (though without striving for detailed biological realism). However, the practical work with DeSTIN to date has focused on visual and auditory processing; and in the context of the present proposal, the intention is to utilize DeSTIN for perception and actuation oriented processing, hybridizing it with CogPrime which will handle abstract cognition and language. Here we will discuss DeSTIN primarily in the perception context, only briefly mentioning the application to actuation which is conceptually similar.

In DeSTIN (see Figure 4.4), perception is carried out by a deep spatiotemporal inference network, which is connected to a similarly architected critic network that provides feedback on the inference network's performance, and an action network that controls actuators based on the activity in the inference network (Figure 4.5 depicts a standard action hierarchy, of which the hierarchy in DeSTIN is an example). The nodes in these networks perform probabilistic pattern recognition according to algorithms to be described below; and the nodes in each of the networks may receive states of nodes in the other networks as inputs, providing rich interconnectivity and synergetic dynamics.

4.3.1.1 Deep versus Shallow Learning for Perceptual Data Processing

The most critical feature of DeSTIN is its uniquely robust approach to modeling the world based on perceptual data. Mimicking the efficiency and robustness by which the human brain analyzes and represents information has been a core challenge in AI research for decades. For instance, humans are exposed to massive amounts of visual and auditory data every second of every day, and are somehow able to capture critical aspects of it in a way that allows for appropriate future recollection and action selection. For decades, it has been known that the

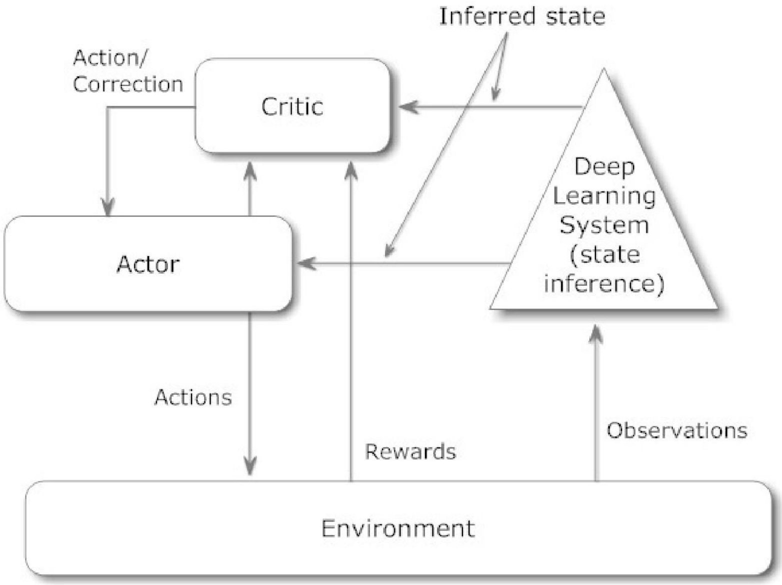


Fig. 4.4: High-level architecture of DeSTIN

brain is a massively parallel fabric, in which computation processes and memory storage are highly distributed. But massive parallelism is not in itself a solution – one also needs the right architecture; which DeSTIN provides, building on prior work in the area of deep learning.

Humanlike intelligence is heavily adapted to the physical environments in which humans evolved; and one key aspect of sensory data coming from our physical environments is its **hierarchical** structure. However, most machine learning and pattern recognition systems are "shallow" in structure, not explicitly incorporating the hierarchical structure of the world in their architecture. In the context of perceptual data processing, the practical result of this is the need to couple each shallow learner with a pre-processing stage, wherein high-dimensional sensory signals are reduced to a lower-dimension feature space that can be understood by the shallow learner. The hierarchical structure of the world is thus crudely captured in the hierarchy of "preprocessor plus shallow learner." In this sort of approach, much of the intelligence of the system shifts to the feature extraction process, which is often imperfect and always application-domain specific.

Deep machine learning has emerged as a more promising framework for dealing with complex, high-dimensional real-world data. Deep learning systems possess a hierarchical structure that intrinsically biases them to recognize the hierarchical patterns present in real-world data. Thus, they hierarchically form a feature space that is driven by regularities in the observations, rather than by hand-crafted techniques. They also offer robustness to many of the distortions and transformations that characterize real-world signals, such as noise, displacement, scaling, etc.

Deep belief networks [HOT06] and Convolutional Neural Networks [LBDE90] have been demonstrated to successfully address pattern inference in high dimensional data (e.g. images). They owe their success to their underlying paradigm of partitioning large data structures into smaller, more manageable units, and discovering the dependencies that may or may not exist

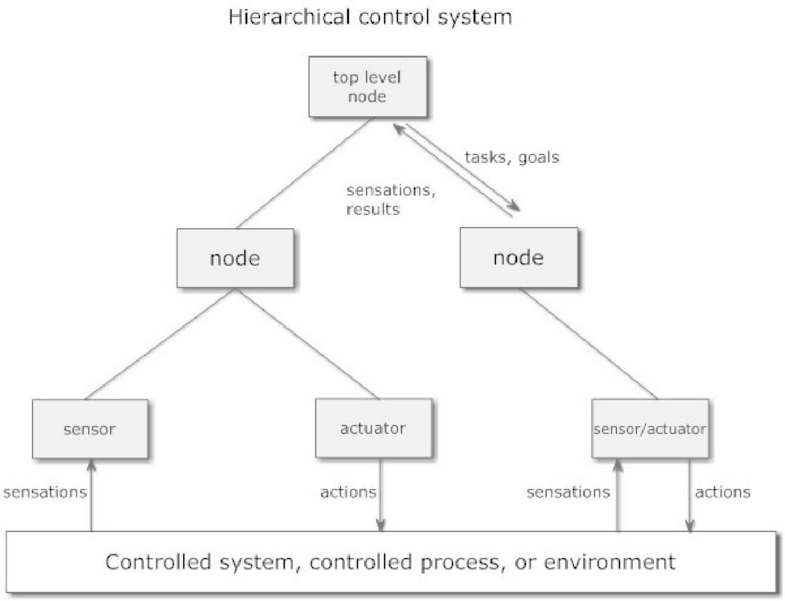

Hierarchical control system
top level node
tasks, goals
sensations, results
node
node
sensor
actuator
sensor/actuator
sensations
actions
sensations
actions
Controlled system, controlled process, or environment

Fig. 4.5: A standard, general-purpose hierarchical control architecture. DeSTIN's control hierarchy exemplifies this architecture, with the difference lying mainly in the DeSTIN control hierarchy's tight integration with the state inference (perception) and critic (reinforcement) hierarchies.

between such units. However, this paradigm has its limitations; for instance, these approaches do not represent temporal information with the same ease as spatial structure. Moreover, some key constraints are imposed on the learning schemes driving these architectures, namely the need for layer-by-layer training, and oftentimes pre-training. DeSTIN overcomes the limitations of prior deep learning approaches to perception processing, and also extends beyond perception to action and reinforcement learning.

4.3.1.2 DeSTIN for Perception Processing

The hierarchical architecture of DeSTIN's spatiotemporal inference network comprises an arrangement into multiple layers of "nodes" comprising multiple instantiations of an identical cortical circuit. Each node corresponds to a particular spatiotemporal region, and uses a statistical learning algorithm to characterize the sequences of patterns that are presented to it by nodes in the layer beneath it. More specifically,

- At the very lowest layer of the hierarchy nodes receive as input raw data (e.g. pixels of an image) and continuously construct a belief state that attempts to characterize the sequences of patterns viewed.

- The second layer, and all those above it, receive as input the belief states of nodes at their corresponding lower layers, and attempt to construct belief states that capture regularities in their inputs.
- Each node also receives as input the belief state of the node above it in the hierarchy (which constitutes "contextual" information)

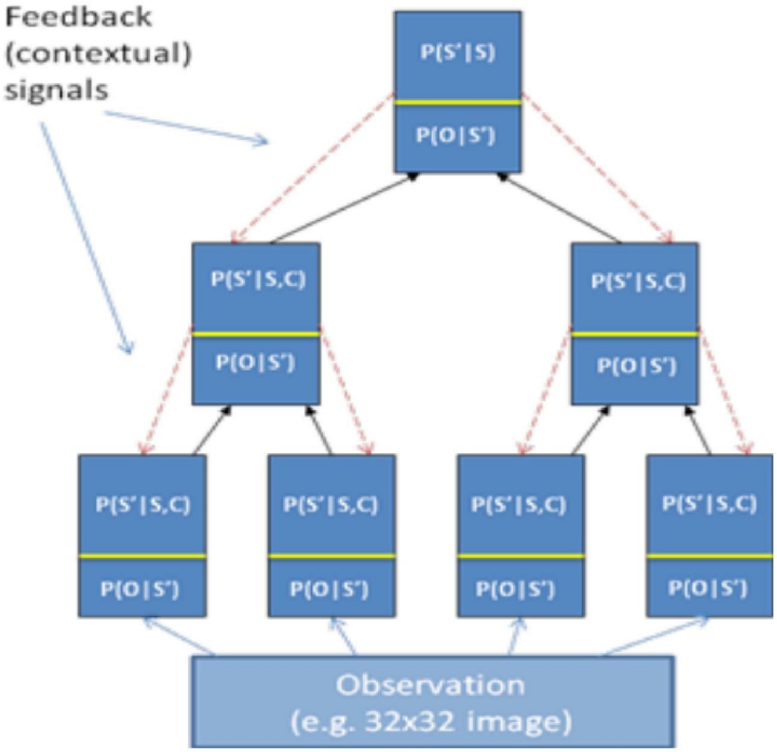


Fig. 4.6: Small-scale instantiation of the DeSTIN perceptual hierarchy. Each box represents a node, which corresponds to a spatiotemporal region (nodes higher in the hierarchy corresponding to larger regions). O denotes the current observation in the region, C is the state of the higher-layer node, and S and S' denote state variables pertaining to two subsequent time steps. In each node, a statistical learning algorithm is used to predict subsequent states based on prior states, current observations, and the state of the higher-layer node.

More specifically, each of the DeSTIN nodes, referring to a specific spacetime region, contains a set of state variables conceived as clusters, each corresponding to a set of previously-observed sequences of events. These clusters are characterized by centroids (and are hence assumed roughly spherical in shape), and each of them comprises a certain "spatiotemporal form" recognized by the system in that region. Each node then contains the task of predicting the likelihood of a certain centroid being most apropos in the near future, based on the past history of observations in the node. This prediction may be done by simple probability tabulation, or via application of supervised learning algorithms such as recurrent neural networks. These clustering and prediction processes occur separately in each node, but the nodes are linked together via bidirectional dynamics: each node feeds input to its parents, and receives "advice" from its parents that is used to condition its probability calculations in a contextual way.

These processes are executed formally by the following basic belief update rule, which governs the learning process and is identical for every node in the architecture. The belief state is a probability mass function over the sequences of stimuli that the nodes learns to represent. Consequently, each node is allocated a predefined number of state variables each denoting a dynamic pattern, or sequence, that is autonomously learned. The DeSTIN update rule maps the current observation (o), belief state (b), and the belief state of a higher-layer node or context (c), to a new (updated) belief state (b'), such that

$$b'(s') = \Pr(s'|o,b,c) = \frac{\Pr(s' \cap o \cap b \cap c)}{\Pr(o \cap b \cap c)}, \tag{4.1}$$

alternatively expressed as

$$b'(s') = \frac{\Pr(o|s',b,c)\Pr(s'|b,c)\Pr(b,c)}{\Pr(o|b,c)\Pr(b,c)}. \tag{4.2}$$

Under the assumption that observations depend only on the true state, or $\Pr(o|s',b,c) = \Pr(o|s')$, we can further simplify the expression such that

$$b'(s') = \frac{\Pr(o|s')\Pr(s'|b,c)}{\Pr(o|b,c)}, \tag{4.3}$$

where $\Pr(s'|b,c) = \sum_{s \in S} \Pr(s'|s,c)\, b(s)$, yielding the belief update rule

$$b'(s') = \frac{\Pr(o|s') \sum_{s \in S} \Pr(s'|s,c)\, b(s)}{\sum_{s'' \in S} \Pr(o|s'') \sum_{s \in S} \Pr(s''|s,c)\, b(s)}, \tag{4.4}$$

where S denotes the sequence set (i.e. belief dimension) such that the denominator term is a normalization factor.

One interpretation of eq. (4.4) would be that the static pattern similarity metric, $\Pr(o|s')$, is modulated by a construct that reflects the system dynamics, $\Pr(s'|s,c)$. As such, the belief state inherently captures both spatial and temporal information. In our implementation, the belief state of the parent node, c, is chosen using the selection rule

$$c = \arg\max_{s} b_p(s), \tag{4.5}$$

where b_p is the belief distribution of the parent node.

A close look at eq. (4.4) reveals that there are two core constructs to be learned, $\Pr(o|s')$ and $\Pr(s'|s,c)$. In the current DeSTIN design, the former is learned via online clustering while the latter is learned based on experience by inductively learning a rule that predicts the next state s' given the prior state s and c.

The overall result is a robust framework that autonomously (i.e. with no human engineered pre-processing of any type) learns to represent complex data patterns, and thus serves the critical role of building and maintaining a model of the state of the world. In a vision processing context, for example, it allows for powerful unsupervised classification. If shown a variety of real-world scenes, it will automatically form internal structures corresponding to the various natural categories of objects shown in the scenes, such as trees, chairs, people, etc.; and also the various natural categories of events it sees, such as reaching, pointing, falling. And, as will be discussed below, it can use feedback from DeSTIN's action and critic networks to further shape its internal world-representation based on reinforcement signals.

Benefits of DeSTIN for Perception Processing

DeSTIN's perceptual network offers multiple key attributes that render it more powerful than other deep machine learning approaches to sensory data processing:

1. The belief space that is formed across the layers of the perceptual network inherently captures both *spatial and temporal regularities* in the data. Given that many applications require that temporal information be discovered for robust inference, this is a key advantage over existing schemes.
2. Spatiotemporal regularities in the observations are captured in a coherent manner (rather than being represented via two separate mechanisms)
3. All processing is both top-down and bottom-up, and both hierarchical and heterarchical, based on nonlinear feedback connections directing activity and modulating learning in multiple directions through DeSTIN's cortical circuits
4. Support for multi-modal fusing is intrinsic within the framework, yielding a powerful state inference system for real-world, partially-observable settings.
5. Each node is identical, which makes it easy to map the design to massively parallel platforms, such as graphics processing units.

Points 2-4 in the above list describe how DeSTIN's perceptual network displays its own "cognitive synergy" in a way that fits naturally into the overall synergetic dynamics of the overall CogPrime architecture. Using this cognitive synergy, DeSTIN's perceptual network addresses a key aspect of general intelligence: the ability to robustly infer the state of the world, with which the system interacts, in an accurate and timely manner.

4.3.1.3 DeSTIN for Action and Control

DeSTIN's perceptual network performs unsupervised world-modeling, which is a critical aspect of intelligence but of course is not the whole story. DeSTIN's action network, coupled with the perceptual network, orchestrates actuator commands into complex movements, but also carries out other functions that are more cognitive in nature.

For instance, people learn to distinguish between cups and bowls in part via hearing other people describe some objects as cups and others as bowls. To emulate this kind of learning, DeSTIN's critic network provides positive or negative reinforcement signals based on whether the action network has correctly identified a given object as a cup or a bowl, and this signal then impacts the nodes in the action network. The critic network takes a simple external "degree of success or failure" signal and turns it into multiple reinforcement signals to be fed into the multiple layers of the action network. The result is that the action network self-organizes so as to include an implicit "cup versus bowl" classifier, whose inputs are the outputs of some of the nodes in the higher levels of the perceptual network. This classifier belongs in the action network because it is part of the procedure by which the DeSTIN system carries out the action of identifying an object as a cup or a bowl.

This example illustrates how the learning of complex concepts and procedures is divided fluidly between the perceptual network, which builds a model of the world in an unsupervised way, and the action network, which learns how to respond to the world in a manner that will receive positive reinforcement from the critic network.

4.3.2 Developmental Robotics Architectures

A particular subset of emergentist cognitive architectures are sufficiently important that we consider them separately here: these are *developmental robotics* architectures, focused on controlling robots without significant "hard-wiring" of knowledge or capabilities, allowing robots to learn (and learn how to learn, etc.) via their engagement with the world. A significant focus is often placed here on "intrinsic motivation," wherein the robot explores the world guided by internal goals like novelty or curiosity, forming a model of the world as it goes along, based on the modeling requirements implied by its goals. Many of the foundations of this research area were laid by Juergen Schmidhuber's work in the 1990s [Sch91b, Sch91a, Sch95, Sch02], but now with more powerful computers and robots the area is leading to more impressive practical demonstrations.

We mention here a handful of the important initiatives in this area:

- Juyang Weng's **Dav** [HZT+02] and **SAIL** [WHZ+00] projects involve mobile robots that explore their environments autonomously, and learn to carry out simple tasks by building up their own world-representations through both unsupervised and teacher-driven processing of high-dimensional sensorimotor data. The underlying philosophy is based on human child development [WH06], the knowledge representations involved are neural network based, and a number of novel learning algorithms are involved, especially in the area of vision processing.
- **FLOWERS** [BO09], an initiative at the French research institute INRIA, led by Pierre-Yves Oudeyer, is also based on a principle of trying to reconstruct the processes of development of the human child's mind, spontaneously driven by intrinsic motivations. Kaplan [Kap08] has taken this project in a direction closely related to our own via the creation of a "robot playroom." Experiential language learning has also been a focus of the project [OK06], driven by innovations in speech understanding.
- **IM-CLEVER**[1], a new European project coordinated by Gianluca Baldassarre and conducted by a large team of researchers at different institutions, is focused on creating software enabling an iCub [MSV+08] humanoid robot to explore the environment and learn to carry out human childlike behaviors based on its own intrinsic motivations. As this project is the closest to our own we will discuss it in more depth below.

Like CogPrime, IM-CLEVER is a humanoid robot intelligence architecture guided by intrinsic motivations, and using hierarchical architectures for reinforcement learning and sensory abstraction. IM-CLEVER's motivational structure is based in part on Schmidhuber's information-theoretic model of curiosity [Sch06]; and CogPrime's Psi-based motivational structure utilizes probabilistic measures of novelty, which are mathematically related to Schmidhuber's measures. On the other hand, IM-CLEVER's use of reinforcement learning follows Schmidhuber's earlier work RL for cognitive robotics [BS04, BZGS06], Barto's work on intrinsically motivated reinforcement learning [SB06, SM05], and Lee's [LMC07b, LMC07a] work on developmental reinforcement learning; whereas CogPrime's assemblage of learning algorithms is more diverse, including probabilistic logic, concept blending and other symbolic methods (in the OCP component) as well as more conventional reinforcement learning methods (in the DeSTIN component).

---

[1] `http://im-clever.noze.it/project/project-description`

In many respects IM-CLEVER bears a moderately strong resemblance to DeSTIN, whose integration with CogPrime is discussed in Chapter 26 of Part 2 (although IM-CLEVER has much more focus on biological realism than DeSTIN). Apart from numerous technical differences, the really big distinction between IM-CLEVER and CogPrime is that in the latter we are proposing to hybridize a hierarchical-abstraction/reinforcement-learning system (such as DeSTIN) with a more abstract symbolic cognition engine that explicitly handles probabilistic logic and language. IM-CLEVER lacks the aspect of hybridization with a symbolic system, taking more of a pure emergentist strategy. Like DeSTIN considered as a standalone architecture IM-CLEVER does entail a high degree of cognitive synergy, between components dealing with perception, world-modeling, action and motivation. However, the "emergentist versus hybrid" is a large qualitative difference between the two approaches.

In all, while we largely agree with the philosophy underlying developmental robotics, our intuition is that the learning and representational mechanisms underlying the current systems in this area are probably not powerful enough to lead to human child level intelligence. We expect that these systems will develop interesting behaviors but fall short of robust preschool level competency, especially in areas like language and reasoning where symbolic systems have typically proved more effective. This intuition is what impels us to pursue a hybrid approach, such as CogPrime. But we do feel that eventually, once the mechanisms underlying brains are better understood and robotic bodies are richer in sensation and more adept in actuation, some sort of emergentist, developmental-robotics approach can be successful at creating humanlike, human-level AGI.

4.4 Hybrid Cognitive Architectures

In response to the complementary strengths and weaknesses of the symbolic and emergentist approaches, in recent years a number of researchers have turned to integrative, hybrid architectures, which combine subsystems operating according to the two different paradigms. The combination may be done in many different ways, e.g. connection of a large symbolic subsystem with a large subsymbolic system, or the creation of a population of small agents each of which is both symbolic and subsymbolic in nature.

Nils Nilsson expressed the motivation for hybrid AGI systems very clearly in his article at the AI-50 conference (which celebrated the 50'th anniversary of the AI field) [Nil09]. While affirming the value of the Physical Symbol System Hypothesis that underlies symbolic AI, he argues that "the PSSH explicitly assumes that, whenever necessary, symbols will be grounded in objects in the environment through the perceptual and effector capabilities of a physical symbol system." Thus, he continues, *"I grant the need for non-symbolic processes in some intelligent systems, but I think they supplement rather than replace symbol systems. I know of no examples of reasoning, understanding language, or generating complex plans that are best understood as being performed by systems using exclusively non-symbolic processes....*

AI systems that achieve human-level intelligence will involve a combination of symbolic and non-symbolic processing."

A few of the more important hybrid cognitive architectures are:

- **CLARION** [SZ04] is a hybrid architecture that combines a symbolic component for reasoning on "explicit knowledge" with a connectionist component for managing "implicit knowledge." Learning of implicit knowledge may be done via neural net, reinforcement learning, or other methods. The integration of symbolic and subsymbolic methods is powerful, but a great deal is still missing such as episodic knowledge and learning and creativity. Learning in the symbolic and subsymbolic portions is carried out separately rather than dynamically coupled, minimizing "cognitive synergy" effects.
- **DUAL** [NK04] is the most impressive system to come out of Marvin Minsky's "Society of Mind" paradigm. It features a population of agents, each of which combines symbolic and connectionist representation, self-organizing to collectively carry out tasks such as perception, analogy and associative memory. The approach seems innovative and promising, but it is unclear how the approach will scale to high-dimensional data or complex reasoning problems due to the lack of a more structured high-level cognitive architecture.
- **LIDA** [BF09] is a comprehensive cognitive architecture heavily based on Bernard Baars' "Global Workspace Theory". It articulates a "cognitive cycle" integrating various forms of memory and intelligent processing in a single processing loop. The architecture ties in well with both neuroscience and cognitive psychology, but it deals most thoroughly with "lower level" aspects of intelligence, handling more advanced aspects like language and reasoning only somewhat sketchily. There is a clear mapping between LIDA structures and processes and corresponding structures and processing in OCP; so that it's only a mild stretch to view CogPrime as an instantiation of the general LIDA approach that extends further both in the lower level (to enable robot action and sensation via DeSTIN) and the higher level (to enable advanced language and reasoning via OCP mechanisms that have no direct LIDA analogues).
- **MicroPsi** [Bac09] is an integrative architecture based on Dietrich Dorner's Psi model of motivation, emotion and intelligence. It has been tested on some practical control applications, and also on simulating artificial agents in a simple virtual world. MicroPsi's comprehensiveness and basis in neuroscience and psychology are impressive, but in the current version of MicroPsi, learning and reasoning are carried out by algorithms that seem unlikely to scale. OCP incorporates the Psi model for motivation and emotion, so that MicroPsi and CogPrime may be considered very closely related systems. But similar to LIDA, MicroPsi currently focuses on the "lower level" aspects of intelligence, not yet directly handling advanced processes like language and abstract reasoning.
- **PolyScheme** [Cas07] integrates multiple methods of representation, reasoning and inference schemes for general problem solving. Each Polyscheme "specialist" models a different aspect of the world using specific representation and inference techniques, interacting with other specialists and learning from them. Polyscheme has been used to model infant reasoning including object identity, events, causality, and spatial relations. The integration of

reasoning methods is powerful, but the overall cognitive architecture is simplistic compared to other systems and seems focused more on problem-solving than on the broader problem of intelligent agent control.

- **Shruti** [SA93] is a fascinating biologically-inspired model of human reflexive inference, which represents in connectionist architecture relations, types, entities and causal rules using focal-clusters. However, much like Hofstadter's earlier Copycat architecture [Hof95], Shruti seems more interesting as a prototype exploration of ideas than as a practical AGI system; at least, after a significant time of development it has not proved significantly effective in any applications
- James Albus's **4D/RCS** robotics architecture shares a great deal with some of the emergentist architectures discussed above, e.g. it has the same hierarchical pattern recognition structure as DeSTIN and HTM, and the same three cross-connected hierarchies as DeSTIN, and shares with the developmental robotics architectures a focus on real-time adaptation to the structure of the world. However, 4D/RCS is not foundationally learning-based but relies on hard-wired architecture and algorithms, intended to mimic the qualitative structure of relevant parts of the brain (and intended to be *augmented* by learning, which differentiates it from emergentist approaches.

As our own CogPrime approach is a hybrid architecture, it will come as no surprise that we believe several of the existing hybrid architectures are fundamentally going in the right direction. However, nearly all the existing hybrid architectures have severe shortcomings which we feel will prevent them from achieving robust humanlike AGI.

Many of the hybrid architectures are in essence "multiple, disparate algorithms carrying out separate functions, encapsulated in black boxes and communicating results with each other." For instance, PolyScheme, ACT-R and CLARION all display this "modularity" property to a significant extent. These architectures lack the rich, real-time interaction between the *internal dynamics* of various memory and learning processes that we believe is critical to achieving humanlike general intelligence using realistic computational resources. On the other hand, those architectures that feature richer integration – such as DUAL, Shruti, LIDA and MicroPsi – have the flaw of relying (at least in their current versions) on overly simplistic learning algorithms, which drastically limits their scalability.

It does seem plausible to us that some of these hybrid architectures could be dramatically extended or modified so as to produce humanlike general intelligence. For instance, one could replace LIDA's learning algorithms with others that interrelate with each other in a nuanced synergetic way; or one could replace MicroPsi's simple learning and reasoning methods with much more powerful and scalable ones acting on the same data structures. However, making these changes would dramatically alter the cognitive architectures in question on multiple levels.

4.4.1 Neural versus Symbolic; Global versus Local

The "symbolic versus emergentist" dichotomy that we have used to structure our review of cognitive architectures is not absolute nor fully precisely defined; it is more of a heuristic distinction. In this section, before plunging into the details of particular hybrid cognitive architectures, we review two other related dichotomies that are useful for understanding hybrid systems: *neural versus symbolic* systems, and *globalist versus localist* knowledge representation.

4.4.1.1 Neural-Symbolic Integration

The distinction between neural and symbolic systems has gotten fuzzier and fuzzier in recent years, with developments such as

- Logic-based systems being used to control embodied agents (hence using logical terms to deal with data that is apparently perception or actuation-oriented in nature, rather than being symbolic in the semiotic sense), see [SS03a] and [GMIH08].
- Hybrid systems combining neural net and logical parts, or using logical or neural net components interchangeably in the same role [LAon].
- Neural net systems being used for strongly symbolic tasks such as automated grammar learning ([Elm91], [Elm91], plus more recent work.)

Figure 4.7 presents a schematic diagram of a generic neural-symbolic system, generalizing from [BH05], a paper that gives an elegant categorization of neural-symbolic AI systems. Figure 4.8 depicts several broad categories of neural-symbolic architecture.

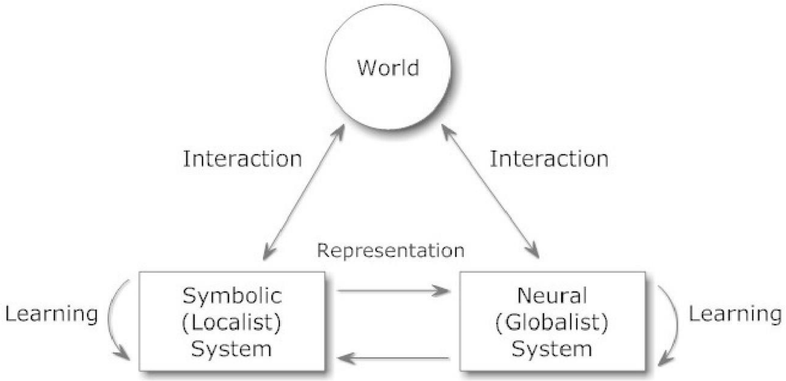


Fig. 4.7: Generic neural-symbolic architecture

Bader and Hitzler categorize neural-symbolic systems according to three orthogonal axes: interrelation, language and usage. “Language” refers to the type of language used in the symbolic component, which may be logical, automata-based, formal grammar-based, etc. “Usage” refers to the purpose to which the neural-symbolic interrelation is put. We tend to use “learning” as an encompassing term for all forms of ongoing knowledge-creation, whereas Bader and Hitzler distinguish learning from reasoning.

Of Bader and Hitzler’s three axes the one that interests us most here is “interrelation”, which refers to the way the neural and symbolic components of the architecture intersect with each other. They distinguish “hybrid” architectures which contain separate but equal, interacting neural and symbolic components; versus “integrative” architectures in which the symbolic component essentially rides piggyback on the neural component, extracting information from it and helping it carry out its learning, but playing a clearly derived and secondary role. We prefer Sun’s (2001) term “monolithic” to Bader and Hitzler’s “integrative” to describe this type of system, as the latter term seems best preserved in its broader meaning.

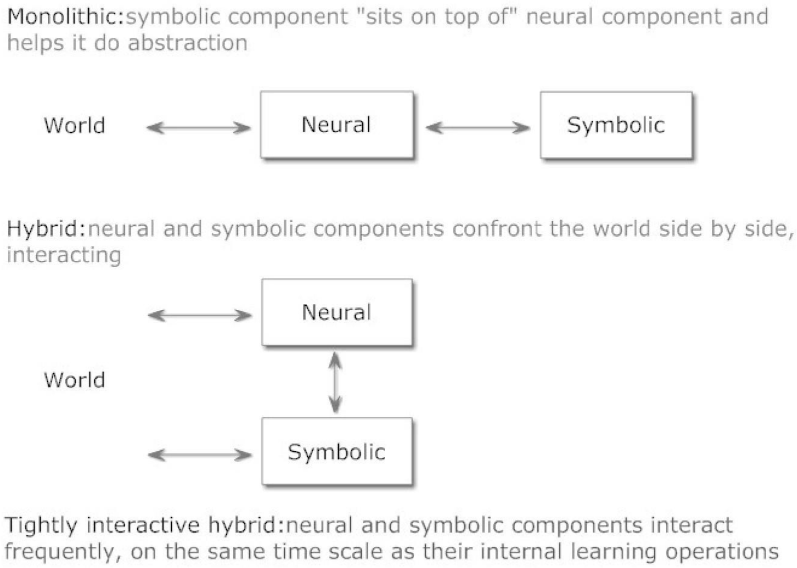

Monolithic:symbolic component "sits on top of" neural component and helps it do abstraction
World
Neural
Symbolic
Hybrid:neural and symbolic components confront the world side by side, interacting
Neural
World
Symbolic
Tightly interactive hybrid:neural and symbolic components interact frequently, on the same time scale as their internal learning operations

Fig. 4.8: Broad categories of neural-symbolic architecture

Within the scope of hybrid neural-symbolic systems, there is another axis which Bader and Hitzler do not focus on, because the main interest of their review is in monolithic systems. We call this axis "interactivity"', and what we are referring to is the frequency of high-information-content, high-influence interaction between the neural and symbolic components in the hybrid system. In a low-interaction hybrid system, the neural and symbolic components don't exchange large amounts of mutually influential information all that frequently, and basically act like independent system components that do their learning/reasoning/thinking periodically sending each other their conclusions. In some cases, interaction may be asymmetric: one component may frequently send a lot of influential information to the other, but not vice versa. However, our hypothesis is that the most capable neural-symbolic systems are going to be the symmetrically highly interactive ones.

In a symmetric high-interaction hybrid neural-symbolic system, the neural and symbolic components exchange influential information sufficiently frequently that each one plays a major role in the other one's learning/reasoning/thinking processes. Thus, the learning processes of each component must be considered as part of the overall dynamic of the hybrid system. The two components aren't just feeding their outputs to each other as inputs, they're mutually guiding each others' internal processing.

One can make a speculative argument for the relevance of this kind of architecture to neuroscience. It seems plausible that this kind of neural-symbolic system roughly emulates the kind of interaction that exists between the brain's neural subsystems implementing localist symbolic processing, and the brain's neural subsystems implementing globalist, classically "connectionist" processing. It seems most likely that, in the brain, symbolic functionality emerges from an underlying layer of neural dynamics. However, it is also reasonable to conjecture that this symbolic functionality is confined to a functionally distinct subsystem of the brain, which then interacts with other subsystems in the brain much in the manner that the symbolic and neural components of a symmetric high-interaction neural-symbolic system interact.

Neuroscience speculations aside, however, our key conjecture regarding neural-symbolic integration is that this sort of neural-symbolic system presents a promising direction for artificial general intelligence research. In Chapter 26 of Volume 2 we will give a more concrete idea of what a symmetric high-interaction hybrid neural-symbolic architecture might look like, exploring the potential for this sort of hybridization between the OpenCogPrime AGI architecture (which is heavily symbolic in nature) and hierarchical attractor neural net based architectures such as DeSTIN.

4.5 Globalist versus Localist Representations

Another interesting distinction, related to but different from "symbolic versus emergentist" and "neural versus symbolic", may be drawn between cognitive systems (or subsystems) where memory is essentially **global**, and those where memory is essentially **local**. In this section we will pursue this distinction in various guises, along with the less familiar notion of **glocal memory**.

This globalist/localist distinction is most easily conceptualized by reference to memories corresponding to categories of entities or events in an external environment. In an AI system that has an internal notion of "activation" – i.e. in which some of its internal elements are more active than others, at any given point in time – one can define the *internal image* of an external event or entity as the fuzzy set of internal elements that tend to be active when that event or entity is presented to the system's sensors. If one has a particular set S of external entities or events of interest, then, the *degree of memory localization* of such an AI system relative to S may be conceived as the percentage of the system's internal elements that have a high degree of membership in the internal image of an average element of S.

Of course, this characterization of localization has its limitations, such as the possibility of ambiguity regarding what are the "system elements" of a given AI system; and the exclusive focus on internal images of external phenomena rather than representation of internal abstract concepts. However, our goal here is not to formulate an ultimate, rigorous and thorough ontology of memory systems, but only to pose a "rough and ready" categorization so as to properly frame our discussion of some specific AGI issues relevant to CogPrime. Clearly the ideas pursued here will benefit from further theoretical exploration and elaboration.

In this sense, a Hopfield neural net [Ami89] would be considered "globalist" since it has a low degree of memory localization (most internal images heavily involve a large number of system elements); whereas Cyc would be considered "localist" as it has a very high degree of memory localization (most internal images are heavily focused on a small set of system elements).

However, although Hopfield nets and Cyc form handy examples, the "globalist vs. localist" distinction as described above is not identical to the "neural vs. symbolic" distinction. For it is in principle quite possible to create localist systems using formal neurons, and also to create globalist systems using formal logic. And "globalist-localist" is not quite identical to "symbolic vs emergentist" either, because the latter is about coordinated system dynamics and behavior not just about knowledge representation. CogPrime combines both symbolic and (loosely) neural representations, and also combines globalist and localist representations in a way that we will call "glocal" and analyze more deeply in Chapter 13; but there are many other ways these various properties could be manifested by AI systems. Rigorously studying the corpus of existing (or hypothetical!) cognitive architectures using these ideas would be a large task, which we do not undertake here.

In the next sections we review several hybrid architectures in more detail, focusing most deeply on LIDA and MicroPsi which have been directly inspirational for CogPrime.

4.5.1 CLARION

Ron Sun's CLARION architecture (see Figure 4.9) is interesting in its combination of symbolic and neural aspects – a combination that is used in a sophisticated way to embody the distinction and interaction between implicit and explicit mental processes. From a CLARION perspective, architectures like Soar and ACT-R are severely limited in that they deal only with explicit knowledge and associated learning processes.

CLARION consists of a number of distinct subsystems, each of which contains a dual representational structure, including a "rules and chunks" symbolic knowledge store somewhat similar to ACT-R, and a neural net knowledge store embodying implicit knowledge. The main subsystems are:

- An action-centered subsystem to control actions;
- A non-action-centered subsystem to maintain general knowledge;
- A motivational subsystem to provide underlying motivations for perception, action, and cognition;
- A meta-cognitive subsystem to monitor, direct, and modify the operations of all the other subsystems.

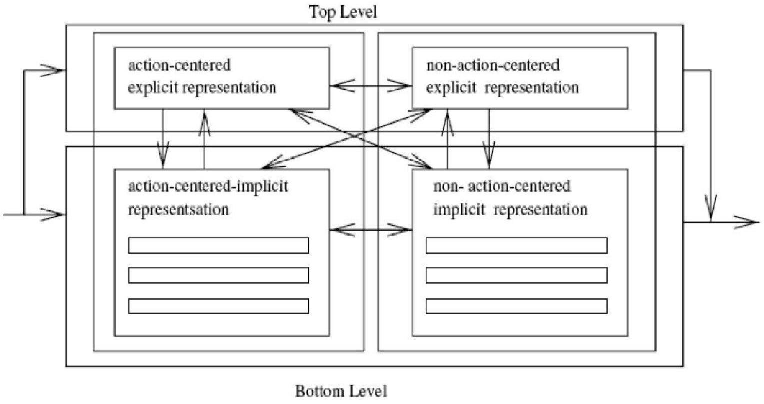

Top Level
action-centered
explicit representation
non-action-centered
explicit representation
action-centered-implicit
represensation
non- action-centered
implicit representation
Bottom Level

Fig. 4.9: The CLARION cognitive architecture.

4.5.2 The Society of Mind and the Emotion Machine

In his influential but controversial book *The Society of Mind* [Min88], Marvin Minsky described a model of human intelligence as something that is built up from the interactions of numerous simple agents. He spells out in great detail how various particular cognitive functions may be achieved via agents and their interactions. He leaves no room for any central algorithms or structures of thought, famously arguing: “What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.”

This perspective was extended in the more recent work *The Emotion Machine* [Min07], where Minsky argued that emotions are “ways to think” evolved to handle different “problem types” that exist in the world. The brain is posited to have rule-based mechanisms (selectors) that turns on emotions to deal with various problems.

Overall, both of these works serve better as works of speculative cognitive science than as works of AI or cognitive architecture per se. As neurologist Richard Restak said in his review of *Emotion Machine*, “Minsky does a marvelous job parsing other complicated mental activities into simpler elements. ... But he is less effective in relating these emotional functions to what’s going on in the brain.” As Restak added, he is also not so effective at relating these emotional functions to straightforwardly implementable algorithms or data structures.

Push Singh, in his PhD thesis and followup work [SBC05], did the best job so far of creating a concrete AI design based on Minsky’s ideas. While Singh’s system was certainly interesting, it was also noteworthy for its lack of any learning mechanisms, and its exclusive focus on explicit rather than implicit knowledge. Due to Singh’s tragic death, his work was never brought anywhere near completion. It seems fair to say that there has not yet been a serious cognitive architecture posed based closely on Minsky’s ideas.

4.5.3 DUAL

The closest thing to a Minsky-ish cognitive architecture is probably DUAL, which takes the Society of Mind concept and adds to it a number of other interesting ideas. DUAL integrates symbolic and connectionist approaches at a deeper level than CLARION, and has been used to model various cognitive functions such as perception, analogy and judgment. Computations in DUAL emerge from the self-organized interaction of many micro-agents, each of which is a hybrid symbolic/connectionist device. Each DUAL agent plays the role of a neural network node, with an activation level and activation spreading dynamics; but also plays the role of a symbol, manipulated using formal rules. The agents exchange messages and activation via links that can be learned and modified, and they form coalitions which collectively represent concepts, episodes, and facts.

The structure of the model is sketchily depicted in Figure 4.10, which covers the application of DUAL to a toy environment called TextWorld. The visual input corresponding to a stimulus is presented on a two-dimensional visual array representing the front end of the system. Perceptual primitives like blobs and terminations are immediately generated by cheap parallel computations. Attention is controlled at each time by an object which allocates it selectively to some area of the stimulus. A detailed symbolic representation is constructed for this area which tends to fade away as attention is withdrawn from it and allocated to another one. Categorization of visual memory contents takes place by retrieving object and scene categories from DUAL's semantic memory and mapping them onto current visual memory representations.

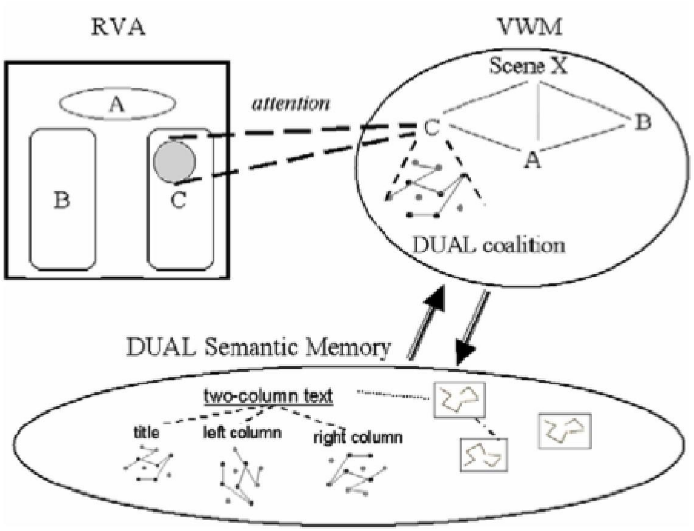


Fig. 4.10: The three main components of the DUAL model: the retinotopic visual array (RVA), the visual working memory (VWM) and DUAL's semantic memory. Attention is allocated to an area of the visual array by the object in VWM controlling attention, while scene and object categories corresponding to the contents of VWM are retrieved from the semantic memory.

In principle the DUAL framework seems quite powerful; using the language of CogPrime, however, it seems to us that the learning mechanisms of DUAL have not been formulated in such a way as to give rise to powerful, scalable cognitive synergy. It would likely be possible to create very powerful AGI systems within DUAL, and perhaps some very CogPrime -like systems as well. But the systems that have been created or designed for use within DUAL so far seem not to be that powerful in their potential or scope.

4.5.4 4D/RCS

In a rather different direction, James Albus, while at the National Bureau of Standards, developed a very thorough and impressive architecture for intelligent robotics called 4D/RCS, which was implemented in a number of machines including unmanned automated vehicles. This architecture lacks critical aspects of intelligence such as learning and creativity, but combines perception, action, planning and world-modeling in a highly effective and tightly-integrated fashion.

The architecture has three hierarchies of memory/processing units: one for perception, one for action and one for modeling and guidance. Each unit has a certain spatiotemporal scope, and (except for the lowest level) supervenes over children whose spatiotemporal scope is a subset of its own. The action hierarchy takes care of decomposing tasks into subtasks; whereas the sensation hierarchy takes care of grouping signals into entities and events. The modeling/guidance hierarchy mediates interactions between perception and action based on its understanding of the world and the system's goals.

In his book [AM01] Albus describes methods for extending 4D/RCS into a complete cognitive architecture, but these extensions have not been elaborated in full detail nor implemented.

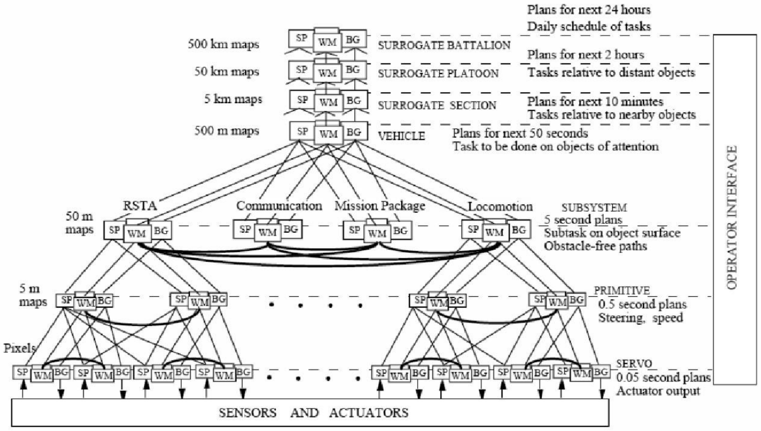

Plans for next 24 hours
Daily schedule of tasks
500 km maps
SP
WM
BG
SURROGATE BATTALION
Plans for next 2 hours
50 km maps
SURROGATE PLATOON
Tasks relative to distant objects
5 km maps
SURROGATE SECTION
Plans for next 10 minutes
Tasks relative to nearby objects
500 m maps
VEHICLE
Plans for next 50 seconds
Task to be done on objects of attention
OPERATOR INTERFACE
RSTA
Communication
Mission Package
Locomotion
50 m maps
SUBSYSTEM
5 second plans
Subtask on object surface
Obstacle-free paths
5 m maps
PRIMITIVE
0.5 second plans
Steering, speed
Pixels
SERVO
0.05 second plans
Actuator output
SENSORS AND ACTUATORS

Fig. 4.11: Albus's 4D-RCS architecture for a single vehicle

4.5.5 PolyScheme

Nick Cassimatis's PolyScheme architecture [Cas07] shares with GLAIR the use of multiple logical reasoning methods on a common knowledge store. While its underlying ideas are quite general, currently PolyScheme is being developed in the context of the "object tracking" domain (construed very broadly). As a logic framework PolyScheme is fairly conventional (unlike GLAIR or NARS with their novel underlying formalisms), but PolyScheme has some unique conceptual aspects, for instance its connection with Cassimatis's theory of mind, which holds that the same core set of logical concepts and relationships underlies both language and physical reasoning [Cas04]. This ties in with the use of a common knowledge store for multiple cognitive processes; for instance it suggests that

- the same core relationships can be used for physical reasoning and parsing, but that each of these domains may involve some additional relationships.
- language processing may be done via physical-reasoning-based cognitive processes, plus the additional activity of some language-specific processes

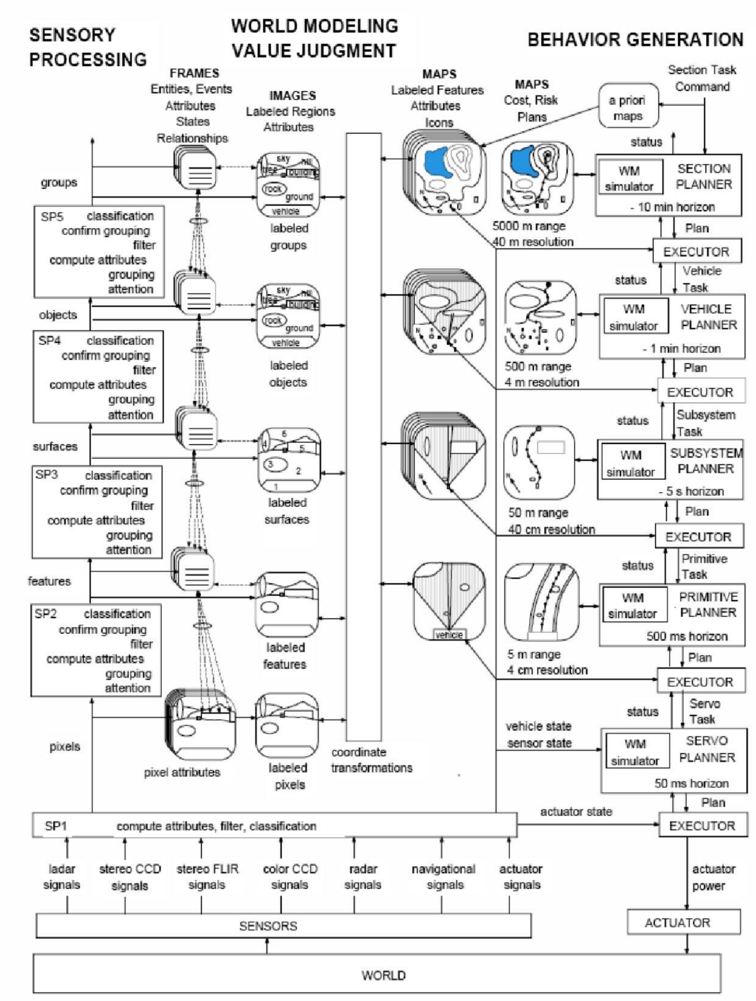

SENSORY PROCESSING
WORLD MODELING VALUE JUDGMENT
BEHAVIOR GENERATION
FRAMES
Entities, Events
Attributes
States
Relationships
IMAGES
Labeled Regions
Attributes
MAPS
Labeled Features
Attributes
Icons
MAPS
Cost, Risk
Plans
a priori maps
Section Task Command
status
groups
SP5 classification confirm grouping filter compute attributes grouping attention
sky
tree
hill
building
rock
ground
vehicle
labeled groups
WM simulator
SECTION PLANNER
- 10 min horizon
5000 m range
40 m resolution
Plan
EXECUTOR
Vehicle Task
objects
SP4 classification confirm grouping filter compute attributes grouping attention
labeled objects
VEHICLE PLANNER
- 1 min horizon
500 m range
4 m resolution
Subsystem Task
surfaces
SP3 classification confirm grouping filter compute attributes grouping attention
labeled surfaces
SUBSYSTEM PLANNER
- 5 s horizon
50 m range
40 cm resolution
Primitive Task
features
SP2 classification confirm grouping filter compute attributes grouping attention
labeled features
vehicle
PRIMITIVE PLANNER
500 ms horizon
5 m range
4 cm resolution
Servo Task
pixels
pixel attributes
labeled pixels
coordinate transformations
vehicle state
sensor state
SERVO PLANNER
50 ms horizon
actuator state
SP1 compute attributes, filter, classification
ladar signals
stereo CCD signals
stereo FLIR signals
color CCD signals
radar signals
navigational signals
actuator signals
actuator power
SENSORS
ACTUATOR
WORLD

Fig. 4.12: Albus's perceptual, motor and modeling hierarchies

4.5.6 Joshua Blue

Sam Adams and his colleagues at IBM have created a cognitive architecture called Joshua Blue [AABL02], which has some significant similarities to CogPrime. Similar to our current research direction with CogPrime, Joshua Blue was created with loose emulation of child cognitive development in mind; and, also similar to CogPrime, it features a number of cognitive processes acting on a common neural-symbolic knowledge store. The specific cognitive processes involved in Joshua Blue and CogPrime are not particularly similar, however. At time of writing (2012)

Joshua Blue is not under active development and has not been for some time; however, the project may be reanimated in future.

Joshua Blue's core knowledge representation is a semantic network of nodes connected by links along which activation spreads. Although many of the nodes have specific semantic referents, as in a classical semantic net, the spread of activation through the network is designed to lead to the emergence of "assemblies" (which could also be thought of as dynamical attractors) in a manner more similar to an attractor neural network.

A major difference from typical semantic or neural network models is the central role that affect plays in the system's dynamics. The weights of the links in the knowledge base are adjusted dynamically based on the emotional context – a very direct way of ensuring that cognitive processes and mental representations are continuously influenced by affect. Qualitatively, this mimics the way that particular emotions in the human brain correlate with the dissemination throughout the brain of particular neurotransmitters, which then affect synaptic activity.

A result of this architecture is that in Joshua Blue, emotion directs attention in a very direct way: affective weighting is important in determining which associated objects will become part of the focus of attention, or will be retained from memory. A notable similarity between CogPrime and Joshua Blue is that in both systems, nodes are assigned two quantitative attention values, one governing allocation of current system resources (mainly processor time; this is CogPrime's ShortTermImportance) and one governing the long-term allocation of memory (CogPrime's LongTermImportance).

The concrete work done with Joshua Blue involved using it to control a simple agent in a simulated world, with the goal that via human interaction, the agent would develop a complex and humanlike emotional and motivational structure from its simple in-built emotions and drives, and would then develop complex cognitive capabilities as part of this development process.

4.5.7 LIDA

The LIDA architecture developed by Stan Franklin and his colleagues [BF09] is based on the concept of the "cognitive cycle" - a notion that is important to nearly every BICA (Biologically Inspired Cognitive Architectures) and also to the brain, but that plays a particularly central role in LIDA. As Franklin says, "as a matter of principle, every autonomous agent, be it human, animal, or artificial, must frequently sample (sense) its environment, process (make sense of) this input, and select an appropriate response (action). The agent's "life" can be viewed as consisting of a continual sequence of iterations of these cognitive cycles. Such cycles constitute the indivisible elements of attention, the least sensing and acting to which we can attend. A cognitive cycle can be thought of as a moment of cognition, a cognitive "moment"."

4.5.8 The Global Workspace

LIDA is heavily based on the "global workspace" concept developed by Bernard Baars. As this concept is also directly relevant to CogPrime it is worth briefly describing here.

In essence Baars' Global Workspace Theory (GWT) is a particular hypothesis about how working memory works and the role it plays in the mind. Baars conceives working memory as the "inner domain in which we can rehearse telephone numbers to ourselves or, more interestingly, in which we carry on the narrative of our lives. It is usually thought to include inner speech and visual imagery." Baars uses the term "consciousness" to refer to the contents of working memory – a theoretical commitment that is not part of the CogPrime design. In this section we will use the term "consciousness" in Baars' way, but not throughout the rest of the book.

Baars conceives working memory and consciousness in terms of a "theater metaphor" – according to which, in the "theater of consciousness" a "spotlight of selective attention" shines a bright spot on stage. The bright spot reveals the global workspace – the contents of consciousness, which may be metaphorically considered as a group of actors moving in and out of consciousness, making speeches or interacting with each other. The unconscious is represented by the audience watching the play ... and there is also a role for the director (the mind's executive processes) behind the scenes, along with a variety of helpers like stage hands, script writers, scene designers, etc.

GWT describes a fleeting memory with a duration of a few seconds. This is much shorter than the 10-30 seconds of classical working memory – according to GWT there is a very brief "cognitive cycle" in which the global workspace is refreshed, and the time period an item remains in working memory generally spans a large number of these elementary "refresh" actions. GWT contents are proposed to correspond to what we are conscious of, and are said to be broadcast to a multitude of unconscious cognitive brain processes. Unconscious processes, operating in parallel, can form coalitions which can act as input processes to the global workspace. Each unconscious process is viewed as relating to certain goals, and seeking to get involved with coalitions that will get enough importance to become part of the global workspace – because once they're in the global workspace they'll be allowed to broadcast out across the mind as a whole, which include broadcasting to the internal and external actuators that allow the mind to do things. Getting into the global workspace is a process's best shot at achieving its goals.

Obviously, the theater metaphor used to describe the GWT is evocative but limited; for instance, the unconscious in the mind does a lot more than the audience in a theater. The unconscious comes up with complex creative ideas sometimes, which feed into consciousness – almost as if the audience is also the scriptwriter. Baars' theory, with its understanding of unconscious dynamics in terms of coalition-building, fails to describe the subtle dynamics occurring within the various forms of long-term memory, which result in subtle nonlinear interactions between long term memory and working memory. But nevertheless, GWT successfully models a number of characteristics of consciousness, including its role in handling novel situations, its limited capacity, its sequential nature, and its ability to trigger a vast range of unconscious brain processes. It is the framework on which LIDA's theory of the cognitive cycle is built.

4.5.9 The LIDA Cognitive Cycle

The simplest cognitive cycle is that of an animal, which senses the world, compares sensation to memory, and chooses an action, all in one fluid subjective moment. But the same cognitive cycle structure/process applies to higher-level cognitive processes as well. The LIDA architecture is based on the LIDA model of the cognitive cycle, which posits a particular structure underlying the cognitive cycle that possess the generality to encompass both simple and complex cognitive moments.

The LIDA cognitive cycle itself is a theoretical construct that can be implemented in many ways, and indeed other BICAs like CogPrime and Psi also manifest the LIDA cognitive cycle in their dynamics, though utilizing different particular structures to do so.

Figure 4.13 shows the cycle pictorially, starting in the upper left corner and proceeding clockwise. At the start of a cycle, the LIDA agent perceives its current situation and allocates attention differentially to various parts of it. It then broadcasts information about the most important parts (which constitute the agent's consciousness), and this information gets features extracted from it, when then get passed along to episodic and semantic memory, that interact in the "global workspace" to create a model for the agent's current situation. This model then, in interaction with procedural memory, enables the agent to choose an appropriate action and execute it - the critical "action-selection" phase!

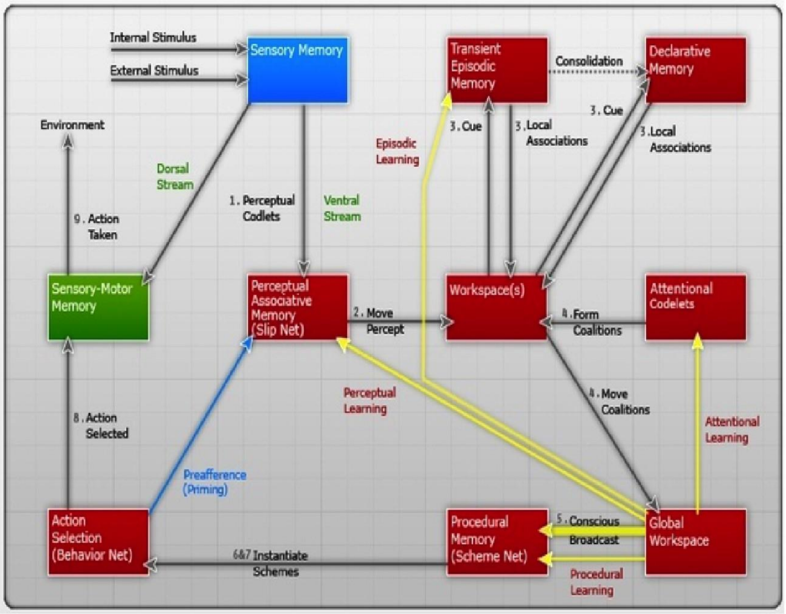

Internal Stimulus
External Stimulus
Sensory Memory
Transient Episodic Memory
Consolidation
Declarative Memory
Environment
Dorsal Stream
1. Perceptual Codlets
Ventral Stream
Episodic Learning
3. Cue
3. Local Associations
3. Cue
3. Local Associations
9. Action Taken
Sensory-Motor Memory
Perceptual Associative Memory (Slip Net)
2. Move Percept
Workspace(s)
4. Form Coalitions
Attentional Codelets
Perceptual Learning
4. Move Coalitions
Attentional Learning
8. Action Selected
Preafference (Priming)
Action Selection (Behavior Net)
6&7 Instantiate Schemes
Procedural Memory (Scheme Net)
5. Conscious Broadcast
Global Workspace
Procedural Learning

Fig. 4.13: The LIDA Cognitive Cycle

The LIDA Cognitive Cycle in More Depth

[2]

We now run through the cognitive cycle in more detail. It begins with sensory stimuli from the agent's external internal environment. Low-level feature detectors in sensory memory begin the process of making sense of the incoming stimuli. These low-level features are passed to perceptual memory where higher-level features, objects, categories, relations, actions, situations, etc. are recognized. These recognized entities, called percepts, are passed to the workspace, where a model of the agent's current situation is assembled.

---

[2] This section paraphrases heavily from [Fra06]

Workspace structures serve as cues to the two forms of episodic memory, yielding both short and long term remembered local associations. In addition to the current percept, the workspace contains recent percepts that haven't yet decayed away, and the agent's model of the then-current situation previously assembled from them. The model of the agent's current situation is updated from the previous model using the remaining percepts and associations. This updating process will typically require looking back to perceptual memory and even to sensory memory, to enable the understanding of relations and situations. This assembled new model constitutes the agent's understanding of its current situation within its world. Via constructing the model, the agent has made sense of the incoming stimuli.

Now attention allocation comes into play, because a real agent lacks the computational resources to work with all parts of its world-model with maximal mental focus. Portions of the model compete for attention. These competing portions take the form of (potentially overlapping) coalitions of structures comprising parts the model. Once one such coalition wins the competition, the agent has decided what to focus its attention on.

And now comes the purpose of all this processing: to help the agent to decide what to do next. The winning coalition passes to the global workspace, the namesake of Global Workspace Theory, from which it is broadcast globally. Though the contents of this conscious broadcast are available globally, the primary recipient is procedural memory, which stores templates of possible actions including their context and possible results.

Procedural memory also stores an activation value for each such template – a value that attempts to measure the likelihood of an action taken within its context producing the expected result. It's worth noting that LIDA makes a rather specific assumption here. LIDA's "activation" values are like the probabilistic truth values of the implications in CogPrime's $Context \wedge Procedure \rightarrow Goal$ triples. However, in CogPrime this probability is not the same as the ShortTermImportance "attention value" associated with the Implication link representing that implication. Here LIDA merges together two concepts that in CogPrime are separate.

Templates whose contexts intersect sufficiently with the contents of the conscious broadcast instantiate copies of themselves with their variables specified to the current situation. These instantiations are passed to the action selection mechanism, which chooses a single action from these instantiations and those remaining from previous cycles. The chosen action then goes to sensorimotor memory, where it picks up the appropriate algorithm by which it is then executed. The action so taken affects the environment, and the cycle is complete.

The LIDA model hypothesizes that all human cognitive processing is via a continuing iteration of such cognitive cycles. It acknowledges that other cognitive processes may also occur, refining and building on the knowledge used in the cognitive cycle (for instance, the cognitive cycle itself doesn't mention abstract reasoning or creativity). But the idea is that these other processes occur in the context of the cognitive cycle, which is the main loop driving the internal and external activities of the organism.

4.5.9.1 Avoiding Combinatorial Explosion via Adaptive Attention Allocation

LIDA avoids combinatorial explosions in its inference processes via two methods, both of which are also important in CogPrime :

- combining reasoning via association with reasoning via deduction

- foundational use of uncertainty in reasoning

One can create an analogy between LIDA's workspace structures and codelets and a logic-based architecture's assertions and functions. However, LIDA's codelets only operate on the structures that are active in the workspace during any given cycle. This includes recent perceptions, their closest matches in other types of memory, and structures recently created by other codelets. The results with the highest estimate of success, i.e. activation, will then be selected.

Uncertainty plays a role in LIDA's reasoning in several ways, most notably through the base activation of its behavior codelets, which depend on the model's estimated probability of the codelet's success if triggered. LIDA observes the results of its behaviors and updates the base activation of the responsible codelets dynamically.

We note that for this kind of uncertain inference/activation interplay to scale well, some level of cognitive synergy must be present; and based on our understanding of LIDA it is not clear to us whether the particular inference and association algorithms used in LIDA possess the requisite synergy.

4.5.9.2 LIDA versus CogPrime

The LIDA cognitive cycle, broadly construed, exists in CogPrime as in other cognitive architectures. To see how, it suffices to map the key LIDA structures into corresponding CogPrime structures, as is done in Table 4.1. Of course this table does not cover all CogPrime processes, as LIDA does not constitute a thorough explanation of CogPrime structure and dynamics. And in most cases the corresponding CogPrime and LIDA processes don't work in exactly the same way; for instance, as noted above, LIDA's action selection relies solely on LIDA's "activation" values, whereas CogPrime's action selection process is more complex, relying on aspects of CogPrime that lack LIDA analogues.

4.5.10 Psi and MicroPsi

We have saved for last the architecture that has the most in common with CogPrime : Joscha Bach's MicroPsi architecture, closely based on Dietrich Dorner's Psi theory. CogPrime has borrowed substantially from Psi in its handling of emotion and motivation; but Psi also has other aspects that differ considerably from CogPrime. Here we will focus more heavily on the points of overlap, but will mention the key points of difference as well.

The overall Psi cognitive architecture, which is centered on the Psi model of the motivational system, is roughly depicted in Figure 4.14.

Psi's motivational system begins with **Demands**, which are the basic factors that motivate the agent. For an animal these would include things like food, water, sex, novelty, socialization, protection of one's children, and so forth. For an intelligent robot they might include things like electrical power, novelty, certainty, socialization, well-being of others and mental growth.

Psi also specifies two fairly abstract demands and posits them as psychologically fundamental (see Figure 4.15):

- **competence**, the effectiveness of the agent at fulfilling its Urges
- **certainty**, the confidence of the agent's knowledge

LIDA	CogPrime
Declarative memory	Atomspace
attentional codelets	Schema that adjust importance of Atoms explicitly
coalitions	maps
global workspace	attentional focus
behavior codelets	schema
procedural memory (scheme net)	procedures in ProcedureRepository; and network of SchemaNodes in the Atomspace
action selection (behavior net)	propagation of STICurrency from goals to actions, and action selection process
transient episodic memory	perceptual atoms entering AT with high STI, which rapidly decreases in most cases
local workspaces	bubbles of interlinked Atoms with moderate importance, focused on by a subset of MindAgents (defined in Chapter 19 of Part 2) for a period of time
perceptual associative memory	HebbianLinks in the AT
sensory memory	spaceserver/timeserver, plus auxiliary stores for other senses
sensorimotor memory	Atoms storing record of actions taken, linked in with Atoms indexed in sensory memory

Table 4.1: CogPrime Analogues of Key LIDA Features

Each demand is assumed to come with a certain “target level” or “target range” (and these may fluctuate over time, or may change as a system matures and develops). An **Urge** is said to develop when a demand deviates from its target range: the urge then seeks to return the demand to its target range. For instance, in an animal-like agent the demand related to food is more clearly described as “fullness,” and there is a target range indicating that the agent is neither too hungry nor too full of food. If the agent’s fullness deviates from this range, an Urge to return the demand to its target range arises. Similarly, if an agent’s novelty deviates from its target range, this means the agent’s life has gotten either too boring or too disconcertingly weird, and the agent gets an Urge for either more interesting activities (in the case of below-range novelty) or more familiar ones (in the case of above-range novelty).

There is also a primitive notion of **Pleasure** (and its opposite, displeasure), which is considered as different from the complex emotion of “happiness.” Pleasure is understood as associated with Urges: pleasure occurs when an Urge is (at least partially) satisfied, whereas displeasure occurs when an urge gets increasingly severe. The degree to which an Urge is satisfied is not necessarily defined instantaneously; it may be defined, for instance, as a time-decaying weighted average of the proximity of the demand to its target range over the recent past.

So, for instance if an agent is bored and gets a lot of novel stimulation, then it experiences some pleasure. If it’s bored and then the monotony of its stimulation gets even more extreme, then it experiences some displeasure.

Note that, according to this relatively simplistic approach, any decrease in the amount of dissatisfaction causes some pleasure; whereas if everything always continues within its acceptable range, there isn’t any pleasure. This may seem a little counterintuitive, but it’s important to understand that these simple definitions of “pleasure” and “displeasure” are not intended to fully capture the natural language concepts associated with those words. The natural language terms are used here simply as heuristics to convey the general character of the processes in-

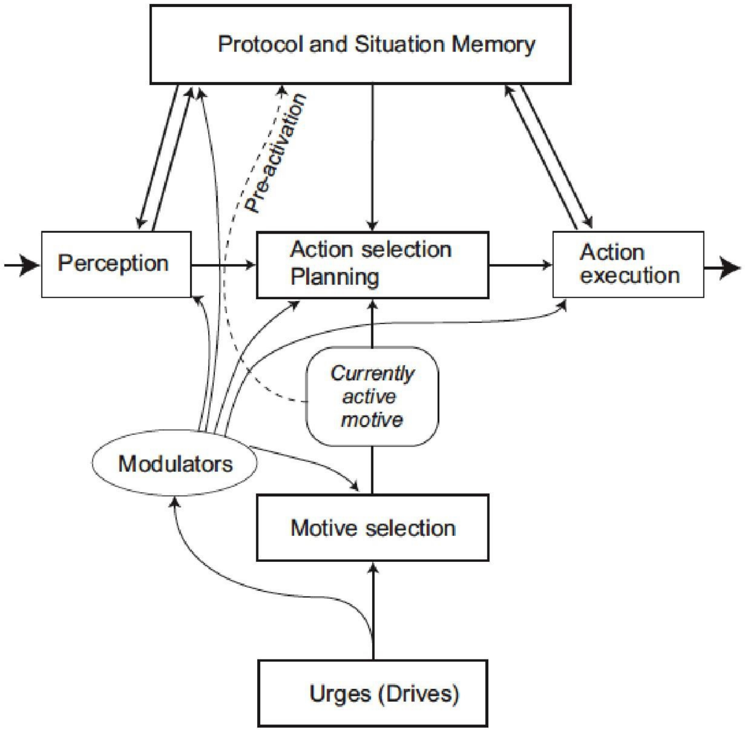

Protocol and Situation Memory
Pre-activation
Perception
Action selection
Planning
Action execution
Currently active motive
Modulators
Motive selection
Urges (Drives)

Fig. 4.14: High-Level Architecture of the Psi Model

volved. These are very low level processes whose analogues in human experience are largely below the conscious level.

A **Goal** is considered as a statement that the system may strive to make true at some future time. A **Motive** is an $(urge, goal)$ pair, consisting of a goal whose satisfaction is predicted to imply the satisfaction of some urge. In fact one may consider Urges as top-level goals, and the agent's other goals as their subgoals.

In Psi an agent has one "ruling motive" at any point in time, but this seems an oversimplification more applicable to simple animals than to human-like or other advanced AI systems. In general one may think of different motives having different weights indicating the amount of resources that will be spent on pursuing them.

Emotions in Psi are considered as complex systemic response-patterns rather than explicitly constructed entities. An emotion is the set of mental entities activated in response to a certain set of urges. Dorner conceived theories about how various common emotions emerge from the dynamics of urges and motives as described in the Psi model. "Intentions" are also considered as composite entities: an intention at a given point in time consists of the active motives, together with their related goals, behavior programs and so forth.

The basic logic of action in Psi is carried out by “triples” that are very similar to CogPrime’s $Context \wedge Procedure \rightarrow Goal$ triples. However, an important role is played by four **modulators** that control how the processes of perception, cognition and action selection are regulated at a given time:

- *activation*, which determines the degree to which the agent is focused on rapid, intensive activity versus reflective, cognitive activity
- *resolution level*, which determines how accurately the system tries to perceive the world
- *certainty*, which determines how hard the system tries to achieve definite, certain knowledge
- *selection threshold*, which determines how willing the system is to change its choice of which goals to focus on

These modulators characterize the system’s emotional and cognitive state at a very abstract level; they are not emotions per se, but they have a large effect on the agent’s emotions. Their intended interaction is depicted in Figure 4.15.

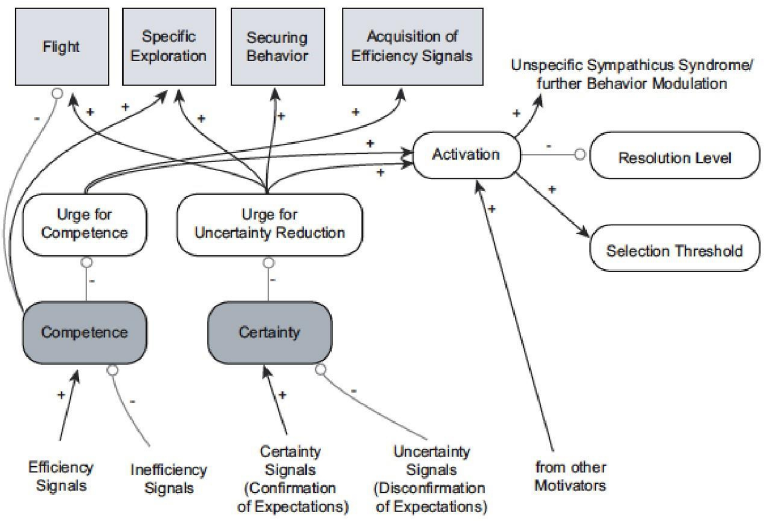

Flight
Specific Exploration
Securing Behavior
Acquisition of Efficiency Signals
Unspecific Sympathicus Syndrome/ further Behavior Modulation
Activation
Resolution Level
Selection Threshold
Urge for Competence
Urge for Uncertainty Reduction
Competence
Certainty
Efficiency Signals
Inefficiency Signals
Certainty Signals (Confirmation of Expectations)
Uncertainty Signals (Disconfirmation of Expectations)
from other Motivators

Fig. 4.15: Primary Interrelationships Between Psi Modulators

4.5.11 The Emergence of Emotion in the Psi Model

We now briefly review the specifics of how Psi models the emergence of emotion. The basic idea is to define a small set of **proto-emotional dimensions** in terms of basic Urges and modulators. Then, emotions are identified with regions in the space spanned by these dimensions.

The simplest approach uses a six-dimensional continuous space:

1. pleasure

2. arousal
3. resolution level
4. selection threshold (i.e. degree of dominance of the leading motive)
5. level of background checks (the rate of the securing behavior)
6. level of goal-directed behavior

Figure 4.16 shows how the latter 5 of these dimensions are derived from underlying urges and modulators. Note that these dimensions are not orthogonal; for instance resolution is mainly inversely related to arousal. Additional dimensions are also discussed, for instance it is postulated that to deal with social emotions one may wish to introduce two more demands corresponding to inner and outer obedience to social norms, and then define dimensions in terms of these.

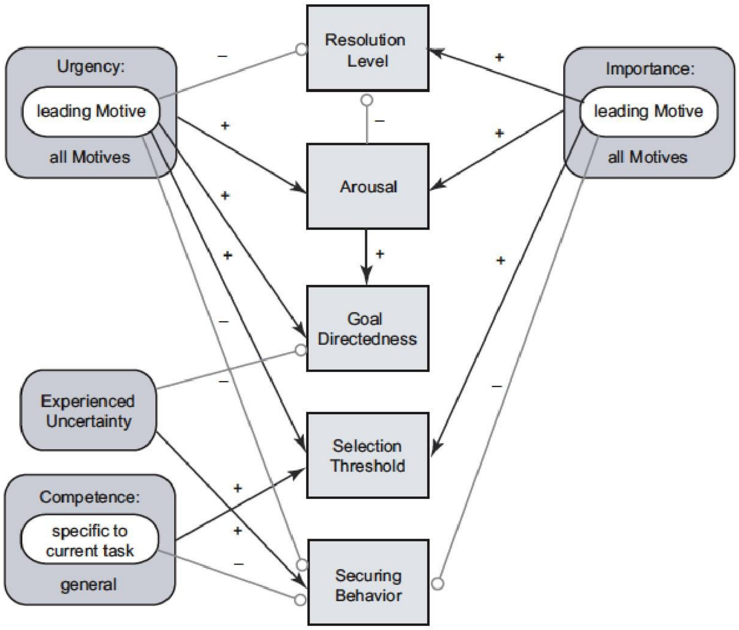

Resolution Level
Urgency:
leading Motive
all Motives
Importance:
leading Motive
all Motives
Arousal
Goal Directedness
Experienced Uncertainty
Selection Threshold
Competence:
specific to current task
general
Securing Behavior
+
–

Fig. 4.16: Five Proto-Emotional Dimensions Implicit in the Psi Model

Specific emotions are then characterized in terms of these dimensions. According to [Bac09], for instance, “Anger ... is characterized by high arousal, low resolution, strong motive dominance, few background checks and strong goal-orientedness; sadness by low arousal, high resolution, strong dominance, few background-checks and low goal-orientedness.”

I’m a bit skeptical of the contention that these dimensions *fully* characterize the relevant emotions. Anger for instance seems to have some particular characteristics not implied by the above list of dimensional values. The list of dimensional values associated with anger doesn’t tell us that an angry person is more likely to punch someone than to bounce up and down, for example. However, it does seem that the dimensional values associated with an emotion are informative about the emotion, so that positioning an emotion on the given dimensions tells one a lot.

4.5.12 Knowledge Representation, Action Selection and Planning in Psi

In addition to the basic motivation/emotion architecture of Psi, which has been adopted (with some minor changes) for use in CogPrime, Psi has a number of other aspects that are somewhat different from their CogPrime analogues.

First of all, on the micro level, Psi represents knowledge using structures called "quads." Each quad is a cluster of 5 neurons containing a core neuron, and four other neurons representing before/after and part-of/has-part relationships in regard to that core neuron. Quads are naturally assembled into spatiotemporal hierarchies, though they are not required to form part of such a structure.

Psi stores knowledge using quads arranged in three networks, which are conceptually similar to the networks in Albus's 4D/RCS and Arel's DeSTIN architectures:

- A sensory network, which stores declarative knowledge: schemas representing images, objects, events and situations as hierarchical structures.
- A motor network, which contains procedural knowledge by way of hierarchical behavior programs
- A motivational network handling demands

Perception in Psi, which is centered in the sensory network, follows principles similar to DeSTIN (which are shared also by other systems), for instance the principle of *perception as prediction.* Psi's "HyPercept" mechanism performs hypothesis-based perception: it attempts to predict what is there to be perceived and then attempts to verify these predictions using sensation and memory. Furthermore HyPercept is intimately coupled with actions in the external world, according to the concept of "Neisser's perceptual cycle," the cycle between exploration and representation of reality. Perceptually acquired information is translated into schemas capable of guiding behaviors, and these are enacted (sometimes affecting the world in significant ways) and in the process used to guide further perception. Imaginary perceptions are handled via a "mental stage" analogous to CogPrime's internal simulation world.

Action selection in Psi works based on what are called "triplets," each of which consists of

- a sensor schema (pre-conditions, "condition schema"; like CogPrime's "context")
- a subsequent motor schema (action, effector; like CogPrime's "procedure")
- a final sensor schema (post-conditions, expectations; like an CogPrime predicate or goal)

What distinguishes these triplets from classic production rules as used in (say) Soar and ACT-R is that the triplets may be partial (some of the three elements may be missing) and may be uncertain. However, there seems no fundamental difference between these triplets and CogPrime's concept/procedure/goal triplets, at a high level; the difference lies in the underlying knowledge representation used for the schemata, and the probabilistic logic used to represent the implication.

The work of figuring out what schema to execute to achieve the chosen goal in the current context is done in Psi using a combination of processes called the "Rasmussen ladder" (named after Danish psychologist Jens Rasmussen). The Rasmussen ladder describes the organization of action as a movement between the stages of skill-based behavior, rule-based behavior and knowledge-based behavior, as follows:

- If a given task amounts to a trained routine, an automatism or skill is activated; it can usually be executed without conscious attention and deliberative control.
- If there is no automatism available, a course of action might be derived from rules; before a known set of strategies can be applied, the situation has to be analyzed and the strategies have to be adapted.
- In those cases where the known strategies are not applicable, a way of combining the available manipulations (operators) into reaching a given goal has to be explored at first. This stage usually requires a recomposition of behaviors, that is, a planning process.

The planning algorithm used in the Psi and MicroPsi implementations is a fairly simple hill-climbing planner. While it's hypothesized that a more complex planner may be needed for advanced intelligence, part of the Psi theory is the hypothesis that most real-life planning an organism needs to do is fairly simple, once the organism has the right perceptual representations and goals.

4.5.13 Psi versus CogPrime

On a high level, the similarities between Psi and CogPrime are quite strong:

- interlinked declarative, procedural and intentional knowledge structures, represented using neural-symbolic methods (though, the knowledge structures have somewhat different high-level structures and low-level representational mechanisms in the two systems)
- perception via prediction and perception/action integration
- action selection via triplets that resemble uncertain, potentially partial production rules
- similar motivation/emotion framework, since CogPrime incorporates a variant of Psi for this

On the nitty-gritty level there are many differences between the systems, but on the big-picture level the *main* difference lies in the way the cognitive synergy principle is pursued in the two different approaches. Psi and MicroPsi rely on very simple learning algorithms that are closely tied to the "quad" neurosymbolic knowledge representation, and hence interoperate in a fairly natural way without need for subtle methods of "synergy engineering." CogPrime uses much more diverse and sophisticated learning algorithms which thus require more sophisticated methods of interoperation in order to achieve cognitive synergy.

Chapter 5
A Generic Architecture of Human-Like Cognition

5.1 Introduction

When writing the first draft of this book, some years ago, we had the idea to explain CogPrime by aligning its various structures and processes with the ones in the "standard architecture diagram" of the human mind. After a bit of investigation, though, we gradually came to the realization that no such thing existed. There was no standard flowchart or other sort of diagram explaining the modern consensus on how human thought works. Many such diagrams existed, but each one seemed to represent some particular focus or theory, rather than an overall integrative understanding.

Since there are multiple opinions regarding nearly every aspect of human intelligence, it would be difficult to get two cognitive scientists to fully agree on every aspect of an overall human cognitive architecture diagram. Prior attempts to outline detailed mind architectures have tended to follow highly specific theories of intelligence, and hence have attracted only moderate interest from researchers not adhering to those theories. An example is Minsky's work presented in *The Emotion Machine* [Min07], which arguably does constitute an architecture diagram for the human mind, but which is only loosely grounded in current empirical knowledge and stands more as a representation of Minsky's own intuitive understanding.

But nevertheless, it seemed to us that a reasonable attempt at an integrative, relatively theory-neutral "human cognitive architecture diagram" would be better than nothing. So naturally, we took it on ourselves to create such a diagram. This chapter is the result – it draws on the thinking of a number of cognitive science and AGI researchers, integrating their perspectives in a coherent, overall architecture diagram for human, and human-like, general intelligence. The specific architecture diagram of CogPrime, given in Chapter 6 below, may then be understood as a particular instantiation of this generic architecture diagram of human-like cognition.

There is no getting around the fact that, to a certain extent, the diagram presented here reflects our particular understanding of how the mind works. However, it was intentionally constructed with the goal of *not* being just an abstracted version of the CogPrime architecture diagram! It does not reflect our own idiosyncratic understanding of human intelligence, as much as a combination of understandings previously presented by multiple researchers (including ourselves), arranged according to our own taste in a manner we find conceptually coherent. With this in mind, we call it the "Integrative Human-Like Cognitive Architecture Diagram," or for short "the integrative diagram." We have made an effort to ensure that as many pieces of the integrative diagram as possible are well grounded in psychological and even neuroscientific data, rather than mainly embodying speculative notions; however, given the current state of knowledge, this could not be done to a complete extent, and there is still some speculation involved here and there.

While based on understandings of human intelligence, the integrative diagram is intended to serve as an architectural outline for human-like general intelligence more broadly. For example, CogPrime is explicitly not intended as a precise emulation of human intelligence, and does many things quite differently than the human mind, yet can still fairly straightforwardly be mapped into the integrative diagram.

The integrative diagram focuses on *structure*, but this should not be taken to represent a valuation of structure over dynamics in our approach to intelligence. Following chapters treat various dynamical phenomena in depth.

5.2 Key Ingredients of the Integrative Human-Like Cognitive Architecture Diagram

The main ingredients we've used in assembling the integrative diagram are as follows:

- Our own views on the various types of memory critical for human-like cognition, and the need for tight, "synergetic" interactions between the cognitive processes focused on these
- Aaron Sloman's high-level architecture diagram of human intelligence [Slo01], drawn from his CogAff architecture, which strikes me as a particularly clear embodiment of "modern common sense" regarding the overall architecture of the human mind. We have added only a couple items to Sloman's high-level diagram, which we felt deserved an explicit high-level role that he did not give them: emotion, language and reinforcement.
- The LIDA architecture diagram presented by Stan Franklin and Bernard Baars [BF09]. We think LIDA is an excellent model of working memory and what Sloman calls "reactive processes", with well-researched grounding in the psychology and neuroscience literature. We have adapted the LIDA diagram only very slightly for use here, changing some of the terminology on the arrows, and indicating where parts of the LIDA diagram indicate processes elaborated in more detail elsewhere in the integrative diagram.
- The architecture diagram of the Psi model of motivated cognition, presented by Joscha Bach in [Bac09] based on prior work by Dietrich Dorner [Dör02]. This diagram is presented without significant modification; however it should be noted that Bach and Dorner present this diagram in the context of larger and richer cognitive models, the other aspects of which are not all incorporated in the integrative diagram.
- James Albus's three-hierarchy model of intelligence [AM01], involving coupled perception, action and reinforcement hierarchies. Albus's model, utilized in the creation of intelligent unmanned automated vehicles, is a crisp embodiment of many ideas emergent from the field of intelligent control systems.
- Deep learning networks as a model of perception (and action and reinforcement learning), as embodied for example in the work of Itamar Arel [ARC09] and Jeff Hawkins [HB06]. The integrative diagram adopts this as the basic model of the perception and action subsystems of human intelligence. Language understanding and generation are also modeled according to this paradigm.

One possible negative reaction to the integrative diagram might be to say that it's a kind of Frankenstein monster diagram, piecing together aspects of different theories in a way that violates the theoretical notions underlying all of them! For example, the integrative diagram takes LIDA as a model of working memory and reactive processing, but from the papers on LIDA it's unclear whether the creators of LIDA construe it more broadly than that. The deep learning community tends to believe that the architecture of current deep learning networks, in itself, is close to sufficient for human-level general intelligence – whereas the integrative diagram appropriates the ideas from this community mainly for handling perception, action and language, etc.

On the other hand, in a more positive perspective, one could view the integrative diagram as consistent with LIDA, but merely providing much more detail on some of the boxes in the LIDA diagram (e.g. dealing with perception and long-term memory). And one could view the integrative diagram as consistent with the deep learning paradigm – via viewing it, not as a description of components to be explicitly implemented in an AGI system, but rather as a description of the key structures and processes that must emerge in deep learning network, based on its engagement with the world, in order for it to achieve human-like general intelligence.

Our own view, underlying the creation of the integrative diagram, is that different communities of cognitive science researchers have focused on different aspects of intelligence, and have thus each created models that are more fully fleshed out in some aspects than others. But these various models all link together fairly cleanly, which is not surprising as they are all grounded in the same data regarding human intelligence. Many judgment calls must be made in fusing multiple models in the way that the integrative diagram does, but we feel these can be made without violating the spirit of the component models. In assembling the integrative diagram, we have made these judgment calls as best we can, but we're well aware that different judgments would also be feasible and defensible. Revisions are likely as time goes on, not only due to new data about human intelligence but also to evolution of understanding regarding the best approach to model integration.

Another possible argument against the ideas presented here is that there's nothing new – all the ingredients presented have been given before elsewhere. To this our retort is to quote Pascal: "Let no one say that I have said nothing new ... the arrangement of the subject is new." The various architecture diagrams incorporated into the integrative diagram are either extremely high level (Sloman's diagram) or focus primarily on one aspect of intelligence, treating the others very concisely by summarizing large networks of distinction structures and processes in small boxes. The integrative diagram seeks to cover all aspects of human-like intelligence at a roughly equal granularity – a different arrangement.

This kind of high-level diagramming exercise is not precise enough, nor dynamics-focused enough, to serve as a guide for creating human-level or more advanced AGI. But it can be a useful tool for explaining and interpreting a concrete AGI design, such as CogPrime.

5.3 An Architecture Diagram for Human-Like General Intelligence

The integrative diagram is presented here in a series of seven Figures.

Figure 5.1 gives a high-level breakdown into components, based on Sloman's high-level cognitive-architectural sketch [Slo01]. This diagram represents, roughly speaking, "modern common sense" about how a human-like mind is architected. The separation between structures

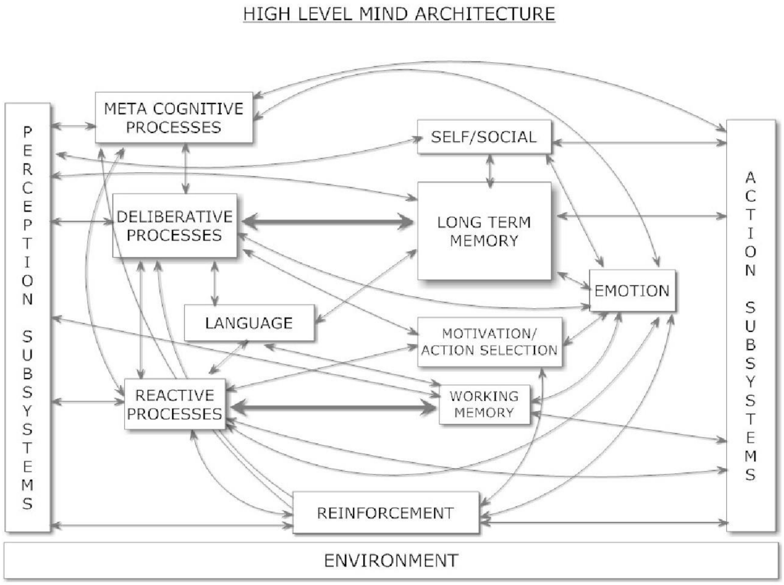

HIGH LEVEL MIND ARCHITECTURE
PERCEPTION SUBSYSTEMS
META COGNITIVE PROCESSES
SELF/SOCIAL
DELIBERATIVE PROCESSES
LONG TERM MEMORY
LANGUAGE
EMOTION
MOTIVATION/ ACTION SELECTION
REACTIVE PROCESSES
WORKING MEMORY
REINFORCEMENT
ACTION SUBSYSTEMS
ENVIRONMENT

Fig. 5.1: High-Level Architecture of a Human-Like Mind

and processes, embodied in having separate boxes for Working Memory vs. Reactive Processes, and for Long Term Memory vs. Deliberative Processes, could be viewed as somewhat artificial, since in the human brain and most AGI architectures, memory and processing are closely integrated. However, the tradition in cognitive psychology is to separate out Working Memory and Long Term Memory from the cognitive processes acting thereupon, so we have adhered to that convention. The other changes from Sloman's diagram are the explicit inclusion of language, representing the hypothesis that language processing is handled in a somewhat special way in the human brain; and the inclusion of a reinforcement component parallel to the perception and action hierarchies, as inspired by intelligent control systems theory (e.g. Albus as mentioned above) and deep learning theory. Of course Sloman's high level diagram in its original form is intended as inclusive of language and reinforcement, but we felt it made sense to give them more emphasis.

Figure 5.2, modeling working memory and reactive processing, is essentially the LIDA diagram as given in prior papers by Stan Franklin, Bernard Baars and colleagues [BF09]. The boxes in the upper left corner of the LIDA diagram pertain to sensory and motor processing, which LIDA does not handle in detail, and which are modeled more carefully by deep learning theory. The bottom left corner box refers to action selection, which in the integrative diagram is modeled in more detail by Psi. The top right corner box refers to Long-Term Memory, which the integrative diagram models in more detail as a synergetic multi-memory system (Figure 5.4).

The original LIDA diagram refers to various "codelets", a key concept in LIDA theory. We have replaced "attention codelets" here with "attention flow", a more generic term. We suggest one can think of an attention codelet as: a piece of information stating that, for a certain group of items, it's currently pertinent to pay attention to this group as a collective.

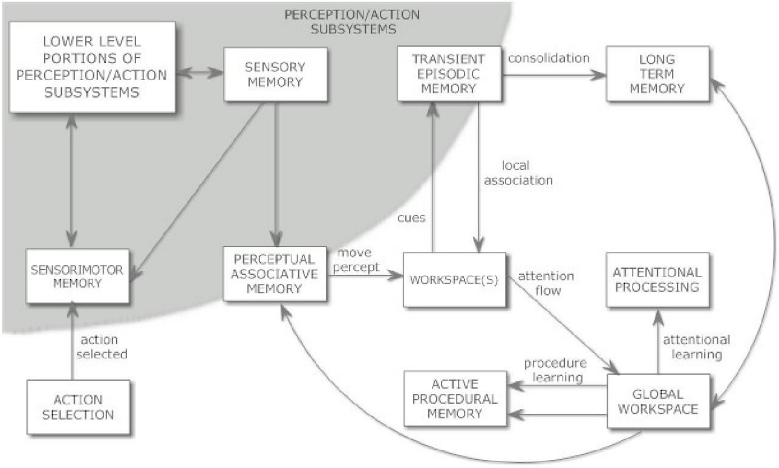

PERCEPTION/ACTION SUBSYSTEMS
LOWER LEVEL PORTIONS OF PERCEPTION/ACTION SUBSYSTEMS
SENSORY MEMORY
TRANSIENT EPISODIC MEMORY
consolidation
LONG TERM MEMORY
local association
cues
SENSORIMOTOR MEMORY
PERCEPTUAL ASSOCIATIVE MEMORY
move percept
WORKSPACE(S)
attention flow
ATTENTIONAL PROCESSING
attentional learning
action selected
procedure learning
ACTIVE PROCEDURAL MEMORY
GLOBAL WORKSPACE
ACTION SELECTION

Fig. 5.2: Architecture of Working Memory and Reactive Processing, closely modeled on the LIDA architecture

Figure 5.3, modeling motivation and action selection, is a lightly modified version of the Psi diagram from Joscha Bach's book *Principles of Synthetic Intelligence* [Bac09]. The main difference from Psi is that in the integrative diagram the Psi motivated action framework is embedded in a larger, more complex cognitive model. Psi comes with its own theory of working and long-term memory, which is related to but different from the one given in the integrative diagram – it views the multiple memory types distinguished in the integrative diagram as emergent from a common memory substrate. Psi comes with its own theory of perception and action, which seems broadly consistent with the deep learning approach incorporated in the integrative diagram. Psi's handling of working memory lacks the detailed, explicit workflow of LIDA, though it seems broadly conceptually consistent with LIDA.

In Figure 5.3, the box labeled "Other portions of working memory" is labeled "Protocol and situation memory" in the original Psi diagram. The Perception, Action Execution and Action Selection boxes have fairly similar semantics to the similarly labeled boxes in the LIDA-like Figure 5.2, so that these diagrams may be viewed as overlapping. The LIDA model doesn't explain action selection and planning in as much detail as Psi, so the Psi-like Figure 5.3 could be viewed as an elaboration of the action-selection portion of the LIDA-like Figure 5.2. In Psi, reinforcement is considered as part of the learning process involved in action selection and planning; in Figure 5.3 an explicit "reinforcement box" has been added to the original Psi diagram, to emphasize this.

Figure 5.4, modeling long-term memory and deliberative processing, is derived from our own prior work studying the "cognitive synergy" between different cognitive processes associated with different types of memory. The division into types of memory is fairly standard. Declarative, procedural, episodic and sensorimotor memory are routinely distinguished; we like to distinguish attentional memory and intentional (goal) memory as well, and view these as the interface between long-term memory and the mind's global control systems. One focus of our AGI design work has been on designing learning algorithms, corresponding to these various types of memory,

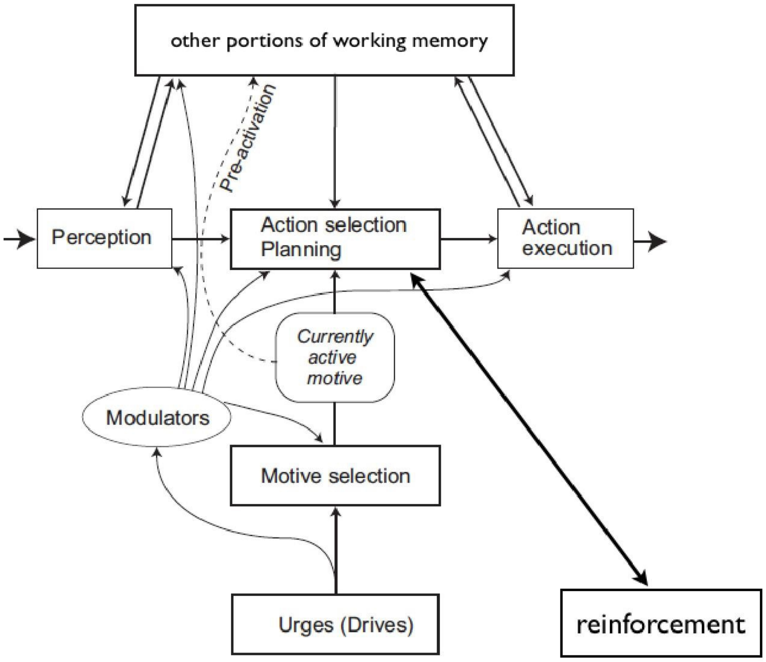

other portions of working memory
Pre-activation
Perception
Action selection
Planning
Action execution
Currently active motive
Modulators
Motive selection
Urges (Drives)
reinforcement

Fig. 5.3: Architecture of Motivated Action

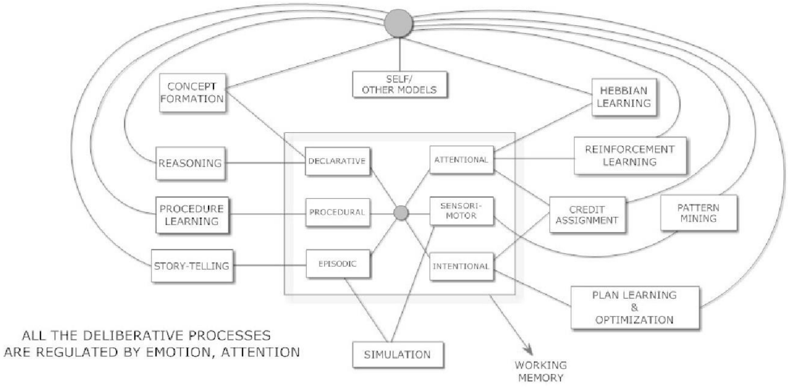

SELF/
OTHER MODELS
CONCEPT FORMATION
HEBBIAN LEARNING
REASONING
DECLARATIVE
ATTENTIONAL
REINFORCEMENT LEARNING
PROCEDURE LEARNING
PROCEDURAL
SENSORI-MOTOR
CREDIT ASSIGNMENT
PATTERN MINING
STORY-TELLING
EPISODIC
INTENTIONAL
PLAN LEARNING & OPTIMIZATION
ALL THE DELIBERATIVE PROCESSES
ARE REGULATED BY EMOTION, ATTENTION
SIMULATION
WORKING MEMORY

Fig. 5.4: Architecture of Long-Term Memory and Deliberative and Metacognitive Thinking

that interact with each other in a synergetic way [Goe09c], helping each other to overcome their intrinsic combinatorial explosions. There is significant evidence that these various types of long-term memory are differently implemented in the brain, but the degree of structure and dynamical commonality underlying these different implementations remains unclear.

Each of these long-term memory types has its analogue in working memory as well. In some cognitive models, the working memory and long-term memory versions of a memory type and corresponding cognitive processes, are basically the same thing. CogPrime is mostly like this – it implements working memory as a subset of long-term memory consisting of items with particularly high importance values. The distinctive nature of working memory is enforced via using slightly different dynamical equations to update the importance values of items with importance above a certain threshold. On the other hand, many cognitive models treat working and long term memory as more distinct than this, and there is evidence for significant functional and anatomical distinctness in the brain in some cases. So for the purpose of the integrative diagram, it seemed best to leave working and long-term memory subcomponents as parallel but distinguished.

Figure 5.4 also encompasses metacognition, under the hypothesis that in human beings and human-like minds, metacognitive thinking is carried out using basically the same processes as plain ordinary deliberative thinking, perhaps with various tweaks optimizing them for thinking about thinking. If it turns out that humans have, say, a special kind of reasoning faculty exclusively for metacognition, then the diagram would need to be modified. Modeling of self and others is understood to occur via a combination of metacognition and deliberative thinking, as well as via implicit adaptation based on reactive processing.

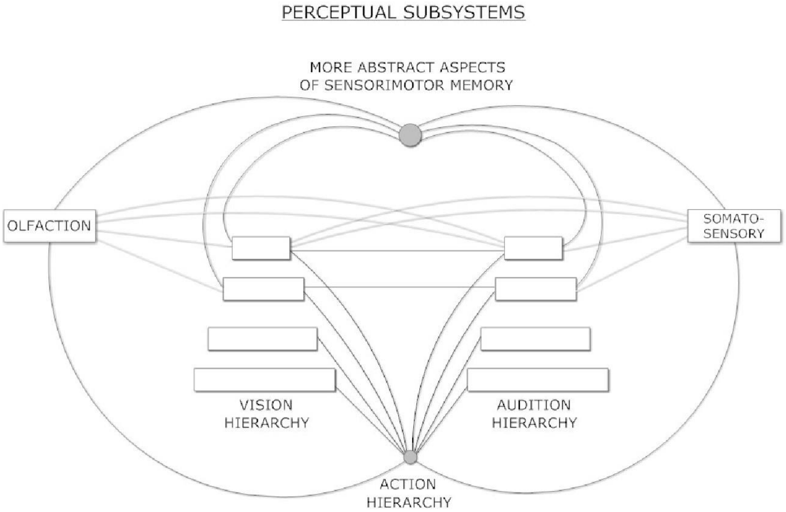


Fig. 5.5: Architecture for Multimodal Perception

Figure 5.5 models perception, according to the basic ideas of deep learning theory. Vision and audition are modeled as deep learning hierarchies, with bottom-up and top-down dynamics. The lower layers in each hierarchy refer to more localized patterns recognized in, and abstracted from, sensory data. Output from these hierarchies to the rest of the mind is not just through the top layers, but via some sort of sampling from various layers, with a bias toward the top layers. The different hierarchies cross-connect, and are hence to an extent dynamically coupled together. It is also recognized that there are some sensory modalities that aren't strongly hierarchical, e.g touch and smell (the latter being better modeled as something like an asymmetric Hopfield net, prone to frequent chaotic dynamics [LLW+05]) – these may also cross-connect with each other and with the more hierarchical perceptual subnetworks. Of course the suggested architecture could include any number of sensory modalities; the diagram is restricted to four just for simplicity.

The self-organized patterns in the upper layers of perceptual hierarchies may become quite complex and may develop advanced cognitive capabilities like episodic memory, reasoning, language learning, etc. A pure deep learning approach to intelligence argues that all the aspects of intelligence emerge from this kind of dynamics (among perceptual, action and reinforcement hierarchies). Our own view is that the heterogeneity of human brain architecture argues against this perspective, and that deep learning systems are probably better as models of perception and action than of general cognition. However, the integrative diagram is not committed to our perspective on this – a deep-learning theorist could accept the integrative diagram, but argue that all the other portions besides the perceptual, action and reinforcement hierarchies should be viewed as descriptions of phenomena that emerge in these hierarchies due to their interaction.

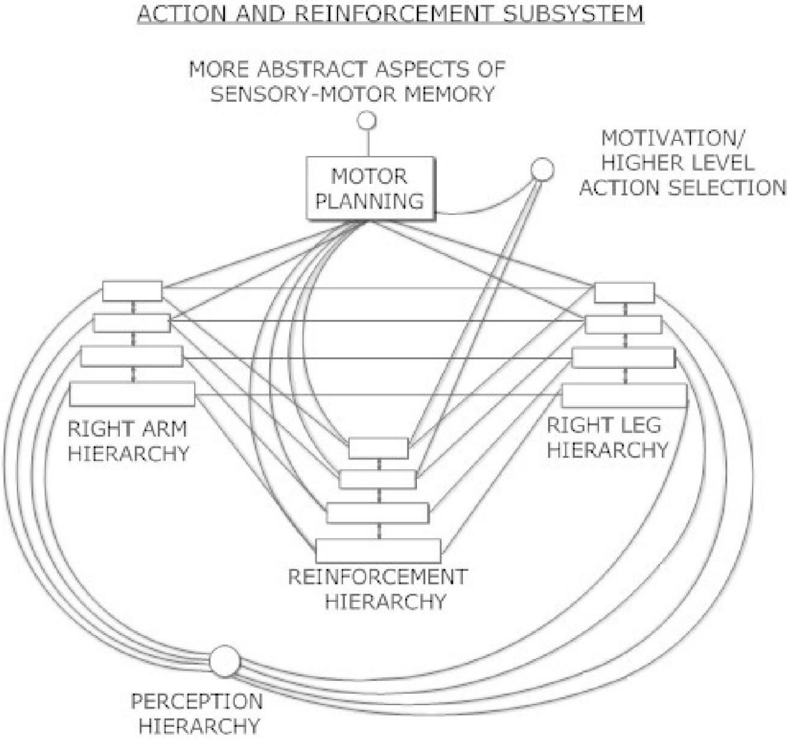

ACTION AND REINFORCEMENT SUBSYSTEM
MORE ABSTRACT ASPECTS OF SENSORY-MOTOR MEMORY
MOTOR PLANNING
MOTIVATION/ HIGHER LEVEL ACTION SELECTION
RIGHT ARM HIERARCHY
RIGHT LEG HIERARCHY
REINFORCEMENT HIERARCHY
PERCEPTION HIERARCHY

Fig. 5.6: Architecture for Action and Reinforcement

Figure 5.6 shows an action subsystem and a reinforcement subsystem, parallel to the perception subsystem. Two action hierarchies, one for an arm and one for a leg, are shown for concreteness, but of course the architecture is intended to be extended more broadly. In the hierarchy corresponding to an arm, for example, the lowest level would contain control patterns corresponding to individual joints, the next level up to groupings of joints (like fingers), the next level up to larger parts of the arm (hand, elbow). The different hierarchies corresponding to different body parts cross-link, enabling coordination among body parts; and they also connect at multiple levels to perception hierarchies, enabling sensorimotor coordination. Finally there is a module for motor planning, which links tightly with all the motor hierarchies, and also overlaps with the more cognitive, inferential planning activities of the mind, in a manner that is modeled different ways by different theorists. Albus [AM01] has elaborated this kind of hierarchy quite elaborately.

The reward hierarchy in Figure 5.6 provides reinforcement to actions at various levels on the hierarchy, and includes dynamics for propagating information about reinforcement up and down the hierarchy.

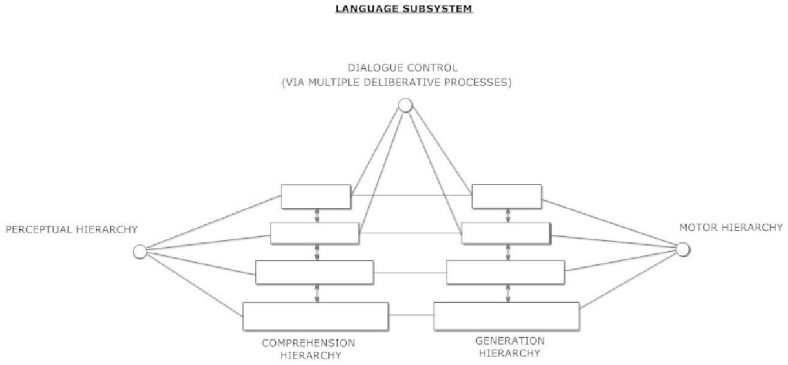


Fig. 5.7: Architecture for Language Processing

Figure 5.7 deals with language, treating it as a special case of coupled perception and action. The traditional architecture of a computational language comprehension system is a pipeline [JM09] [Goe10d], which is equivalent to a hierarchy with the lowest-level linguistic features (e.g. sounds, words) at the bottom, and the highest level features (semantic abstractions) at the top, and syntactic features in the middle. Feedback connections enable semantic and cognitive modulation of lower-level linguistic processing. Similarly, language generation is commonly modeled hierarchically, with the top levels being the ideas needing verbalization, and the bottom level corresponding to the actual sentence produced. In generation the primary flow is top-down, with bottom-up flow providing modulation of abstract concepts by linguistic surface forms.

So, that's it – an integrative architecture diagram for human-like general intelligence, split among seven different pictures, formed by judiciously merging together architecture diagrams produced via a number of cognitive theorists with different, overlapping foci and research paradigms.

Is anything critical left out of the diagram? A quick perusal of the table of contents of cognitive psychology textbooks suggests to me that if anything major is left out, it's also unknown to current cognitive psychology. However, one could certainly make an argument for explicit inclusion of certain other aspects of intelligence, that in the integrative diagram are left as implicit emergent phenomena. For instance, creativity is obviously very important to intelligence, but, there is no "creativity" box in any of these diagrams – because in our view, and the view of the cognitive theorists whose work we've directly drawn on here, creativity is best viewed as a process emergent from other processes that are explicitly included in the diagrams.

5.4 Interpretation and Application of the Integrative Diagram

A tongue-partly-in-cheek definition of a biological pathway is "a subnetwork of a biological network, that fits on a single journal page." Cognitive architecture diagrams have a similar property – they are crude abstractions of complex structures and dynamics, sculpted in accordance with the size of the printed page, and the tolerance of the human eye for absorbing diagrams, and the tolerance of the human author for making diagrams.

However, sometimes constraints – even arbitrary ones – are useful for guiding creative efforts, due to the fact that they force choices. Creating an architecture for human-like general intelligence that fits in a few (okay, seven) fairly compact diagrams, requires one to make many choices about what features and relationships are most essential. In constructing the integrative diagram, we have sought to make these choices, not purely according to our own tastes in cognitive theory or AGI system design, but according to a sort of blend of the taste and judgment of a number of scientists whose views we respect, and who seem to have fairly compatible, complementary perspectives.

What is the use of a cognitive architecture diagram like this? It can help to give newcomers to the field a basic idea about what is known and suspected about the nature of human-like general intelligence. Also, it could potentially be used as a tool for cross-correlating different AGI architectures. If everyone who authored an AGI architecture would explain how their architecture accounts for each of the structures and processes identified in the integrative diagram, this would give a means of relating the various AGI designs to each other.

The integrative diagram could also be used to help connect AGI and cognitive psychology to neuroscience in a more systematic way. In the case of LIDA, a fairly careful correspondence has been drawn up between the LIDA diagram nodes and links and various neural structures and processes [FB08]. Similar knowledge exists for the rest of the integrative diagram, though not organized in such a systematic fashion. A systematic curation of links between the nodes and links in the integrative diagram and current neuroscience knowledge, would constitute an interesting first approximation of the holistic cognitive behavior of the human brain.

Finally (and harking forward to later chapters), the big omission in the integrative diagram is *dynamics*. Structure alone will only get you so far, and you could build an AGI system with reasonable-looking things in each of the integrative diagram's boxes, interrelating according to the given arrows, and yet still fail to make a viable AGI system. Given the limitations the real world places on computing resources, it's not enough to have adequate representations and algorithms in all the boxes, communicating together properly and capable doing the right things given sufficient resources. Rather, one needs to have all the boxes filled in properly with structures and processes that, when they act together using feasible computing resources, will yield appropriately intelligent behaviors via their cooperative activity. And this has to do with the complex interactive dynamics of all the processes in all the different boxes – which is something the integrative diagram doesn't touch at all. This brings us again to the network of ideas we've discussed under the name of "cognitive synergy," to be discussed later on.

It might be possible to make something similar to the integrative diagram on the level of dynamics rather than structures, complementing the structural integrative diagram given here; but this would seem significantly more challenging, because we lack a standard set of tools for depicting system dynamics. Most cognitive theorists and AGI architects describe their structural ideas using boxes-and-lines diagrams of some sort, but there is no standard method for depicting complex system dynamics. So to make a dynamical analogue to the integrative diagram, via a similar integrative methodology, one would first need to create appropriate diagrammatic formalizations of the dynamics of the various cognitive theories being integrated – a fascinating but onerous task.

When we first set out to make an integrated cognitive architecture diagram, via combining the complementary insights of various cognitive science and AGI theorists, we weren't sure how well it would work. But now we feel the experiment was generally a success – the resultant integrated architecture seems sensible and coherent, and reasonably complete. It doesn't come close to telling you everything you need to know to understand or implement a human-like mind – but it tells you the various processes and structures you need to deal with, and which of their interrelations are most critical. And, perhaps just as importantly, it gives a concrete way of understanding the insights of a specific but fairly diverse set of cognitive science and AGI theorists as complementary rather than contradictory. In a CogPrime context, it provides a way of tying in the specific structures and dynamics involved in CogPrime, with a more generic portrayal of the structures and dynamics of human-like intelligence.

Chapter 6
A Brief Overview of CogPrime

6.1 Introduction

Just as there are many different approaches to human flight – airplanes, helicopters, balloons, spacecraft, and doubtless many methods no person has thought of yet – similarly, there are likely many different approaches to advanced artificial general intelligence. All the different approaches to flight exploit the same core principles of aerodynamics in different ways; and similarly, the various different approaches to AGI will exploit the same core principles of general intelligence in different ways.

In the chapters leading up to this one, we have taken a fairly broad view of the project of engineering AGI. We have presented a conception and formal model of intelligence, and described environments, teaching methodologies and cognitive and developmental pathways that we believe are collectively appropriate for the creation of AGI at the human level and ultimately beyond, and with a roughly human-like bias to its intelligence. These ideas stand alone and may be compatible with a variety of approaches to engineering AGI systems. However, they also set the stage for the presentation of CogPrime, the particular AGI design on which we are currently working.

The thorough presentation of the CogPrime design is the job of Part 2 of this book – where, not only are the algorithms and structures involved in CogPrime reviewed in more detailed, but their relationship to the theoretical ideas underlying CogPrime is pursued more deeply. The job of this chapter is a smaller one: to give a high-level overview of some key aspects the CogPrime architecture at a mostly nontechnical level, so as to enable you to approach Part 2 with a little more idea of what to expect. The remainder of Part 1, following this chapter, will present various theoretical notions enabling the particulars, intent and consequences of the CogPrime design to be more thoroughly understood.

6.2 High-Level Architecture of CogPrime

Figures 6.1, 6.2 , 6.4 and 6.5 depict the high-level architecture of CogPrime, which involves the use of multiple cognitive processes associated with multiple types of memory to enable an intelligent agent to execute the procedures that it believes have the best probability of working toward its goals in its current context. In a robot preschool context, for example, the top-level goals will be simple things such as pleasing the teacher, learning new information and skills, and protecting the robot's body. Figure 6.3 shows part of the architecture via which cognitive processes interact with each other, via commonly acting on the AtomSpace knowledge repository.

Comparing these diagrams to the integrative human cognitive architecture diagrams given in Chapter 5, one sees the main difference is that the CogPrime diagrams commit to specific structures (e.g. knowledge representations) and processes, whereas the generic integrative architecture diagram refers merely to types of structures and processes. For instance, the integrative diagram refers generally to declarative knowledge and learning, whereas the CogPrime diagram refers to PLN, as a specific system for reasoning and learning about declarative knowledge. Table 6.1 articulates the key connections between the components of the CogPrime diagram and those of the integrative diagram, thus indicating the general cognitive functions instantiated by each of the CogPrime components.

6.3 Current and Prior Applications of OpenCog

Before digging deeper into the theory, and elaborating some of the dynamics underlying the above diagrams, we pause to briefly discuss some of the practicalities of work done with the OpenCog system currently implementing parts of the CogPrime architecture.

OpenCog, the open-source software framework underlying the "OpenCogPrime" (currently partial) implementation of the CogPrime architecture, has been used for commercial applications in the area of natural language processing and data mining; for instance, see [GPPG06] where OpenCogPrime's PLN reasoning and RelEx language processing are combined to do automated biological hypothesis generation based on information gathered from PubMed abstracts. Most relevantly to the present work, it has also been used to control virtual agents in virtual worlds [GEA08].

Prototype work done during 2007-2008 involved using an OpenCog variant called the OpenPetBrain to control virtual dogs in a virtual world (see Figure 6.6 for a screenshot of an OpenPetBrain-controlled virtual dog). While these OpenCog virtual dogs did not display intelligence closely comparable to that of real dogs (or human children), they did demonstrate a variety of interesting and relevant functionalities including:

- learning new behaviors based on imitation and reinforcement
- responding to natural language commands and questions, with appropriate actions and natural language replies
- spontaneous exploration of their world, remembering their experiences and using them to bias future learning and linguistic interaction

One current OpenCog initiative involves extending the virtual dog work via using OpenCog to control virtual agents in a game world inspired by the game Minecraft. These agents are initially specifically concerned with achieving goals in a game world via constructing structures with blocks and carrying out simple English communications. Representative example tasks would be:

- Learning to build steps or ladders to get desired objects that are high up
- Learning to build a shelter to protect itself from aggressors

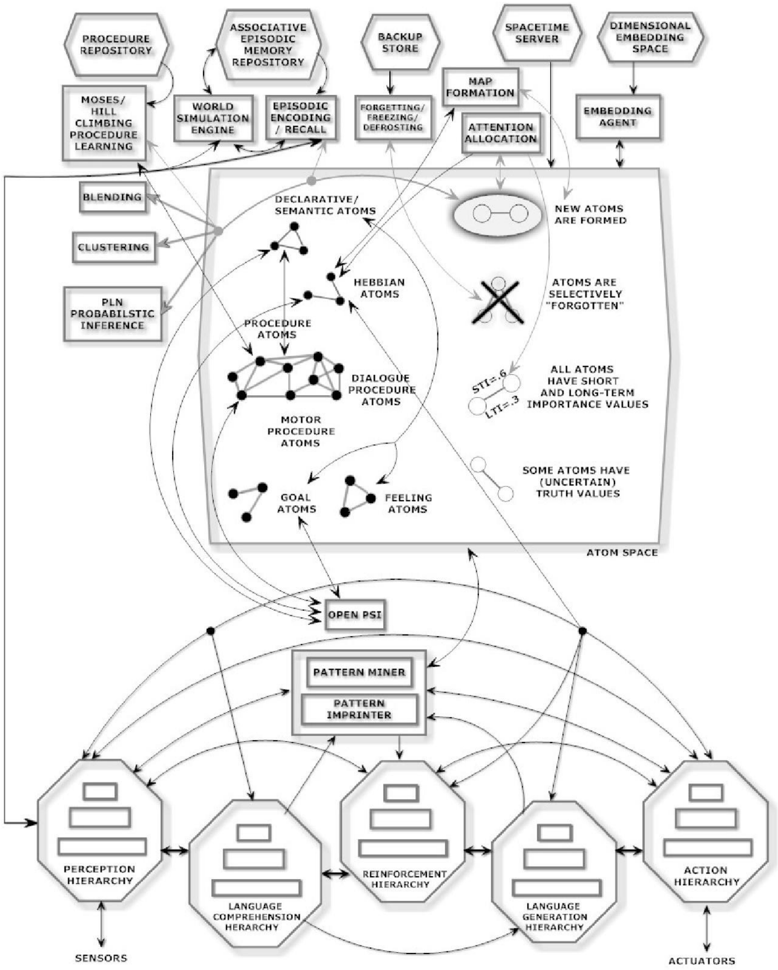

PROCEDURE REPOSITORY
ASSOCIATIVE EPISODIC MEMORY REPOSITORY
BACKUP STORE
SPACETIME SERVER
DIMENSIONAL EMBEDDING SPACE
MOSES/ HILL CLIMBING PROCEDURE LEARNING
WORLD SIMULATION ENGINE
EPISODIC ENCODING / RECALL
FORGETTING/ FREEZING/ DEFROSTING
MAP FORMATION
ATTENTION ALLOCATION
EMBEDDING AGENT
BLENDING
CLUSTERING
PLN PROBABILSTIC INFERENCE
DECLARATIVE/ SEMANTIC ATOMS
NEW ATOMS ARE FORMED
HEBBIAN ATOMS
ATOMS ARE SELECTIVELY "FORGOTTEN"
PROCEDURE ATOMS
DIALOGUE PROCEDURE ATOMS
STI=.6
LTI=.3
ALL ATOMS HAVE SHORT AND LONG-TERM IMPORTANCE VALUES
MOTOR PROCEDURE ATOMS
SOME ATOMS HAVE (UNCERTAIN) TRUTH VALUES
GOAL ATOMS
FEELING ATOMS
ATOM SPACE
OPEN PSI
PATTERN MINER
PATTERN IMPRINTER
PERCEPTION HIERARCHY
LANGUAGE COMPREHENSION HERARCHY
REINFORCEMENT HIERARCHY
LANGUAGE GENERATION HIERARCHY
ACTION HIERARCHY
SENSORS
ACTUATORS

Fig. 6.1: **High-Level Architecture of CogPrime.** This is a conceptual depiction, not a detailed flowchart (which would be too complex for a single image). Figures 6.2 , 6.4 and 6.5 highlight specific aspects of this diagram.

- Learning to build structures resembling structures that it's shown (even if the available materials are a bit different)
- Learning how to build bridges to cross chasms

Of course, the AI significance of learning tasks like this all depends on what kind of feedback the system is given, and how complex its environment is. It would be relatively simple to make an AI system do things like this in a trivial and highly specialized way, but that is not the intent of the project the goal is to have the system learn to carry out tasks like this using general learning mechanisms and a general cognitive architecture, based on embodied experience and only scant feedback from human teachers. If successful, this will provide an outstanding platform for ongoing AGI development, as well as a visually appealing and immediately meaningful demo for OpenCog.

Specific, particularly simple tasks that are the focus of this project team's current work at time of writing include:

- Watch another character build steps to reach a high-up object
- Figure out via imitation of this that, in a different context, building steps to reach a high up object may be a good idea
- Also figure out that, if it wants a certain high-up object but there are no materials for building steps available, finding some other way to get elevated will be a good idea that may help it get the object

6.3.1 Transitioning from Virtual Agents to a Physical Robot

Preliminary experiments have also been conducted using OpenCog to control a Nao robot as well as a virtual dog [GdG08]. This involves hybridizing OpenCog with a separate (but interlinked) subsystem handling low-level perception and action. In the experiments done so far, this has been accomplished in an extremely simplistic way. How to do this right is a topic treated in detail in Chapter 26 of Part 2.

We suspect that reasonable level of capability will be achievable by simply interposing DeSTIN (or some other system in its place) as a perception/action "black box" between OpenCog and a robot. Some preliminary experiments in this direction have already been carried out, connecting the OpenPetBrain to a Nao robot using simpler, less capable software than DeSTIN in the intermediary role (off-the-shelf speech-to-text, text-to-speech and visual object recognition software).

However, we also suspect that to achieve robustly intelligent robotics we must go beyond this approach, and connect robot perception and actuation software with OpenCogPrime in a "white box" manner that allows intimate dynamic feedback between perceptual, motoric, cognitive and linguistic functions. We will achieve this via the creation and real-time utilization of links between the nodes in CogPrime's and DeSTIN's internal networks (a topic to be explored in more depth later in this chapter).

6.4 Memory Types and Associated Cognitive Processes in CogPrime

Now we return to the basic description of the CogPrime approach, turning to aspects of the relationship between structure and dynamics. Architecture diagrams are all very well, but, ultimately it is dynamics that makes an architecture come alive. Intelligence is all about learning, which is by definition about change, about dynamical response to the environment and internal self-organizing dynamics.

CogPrime relies on multiple memory types and, as discussed above, is founded on the premise that the right course in architecting a pragmatic, roughly human-like AGI system is to handle different types of memory differently in terms of both structure and dynamics.

CogPrime's memory types are the declarative, procedural, sensory, and episodic memory types that are widely discussed in cognitive neuroscience [TC05], plus attentional memory for allocating system resources generically, and intentional memory for allocating system resources in a goal-directed way. Table 6.2 overviews these memory types, giving key references and indicating the corresponding cognitive processes, and also indicating which of the generic patternist cognitive dynamics each cognitive process corresponds to (pattern creation, association, etc.). Figure 6.7 illustrates the relationships between several of the key memory types in the context of a simple situation involving an OpenCogPrime-controlled agent in a virtual world.

In terms of patternist cognitive theory, the multiple types of memory in CogPrime should be considered as specialized ways of storing particular types of patterns, optimized for spacetime efficiency. The cognitive processes associated with a certain type of memory deal with creating and recognizing patterns of the type for which the memory is specialized. While in principle all the different sorts of pattern could be handled in a unified memory and processing architecture, the sort of specialization used in CogPrime is necessary in order to achieve acceptable efficient general intelligence using currently available computational resources. And as we have argued in detail in Chapter 7, efficiency is not a side-issue but rather the essence of real-world AGI (since as Hutter has shown, if one casts efficiency aside, arbitrary levels of general intelligence can be achieved via a trivially simple program).

The essence of the CogPrime design lies in the way the structures and processes associated with each type of memory are designed to work together in a closely coupled way, yielding cooperative intelligence going beyond what could be achieved by an architecture merely containing the same structures and processes in separate "black boxes."

The inter-cognitive-process interactions in OpenCog are designed so that

- conversion between different types of memory is possible, though sometimes computationally costly (e.g. an item of declarative knowledge may with some effort be interpreted procedurally or episodically, etc.)
- when a learning process concerned centrally with one type of memory encounters a situation where it learns very slowly, it can often resolve the issue by converting some of the relevant knowledge into a different type of memory: i.e. **cognitive synergy**

6.4.1 Cognitive Synergy in PLN

To put a little meat on the bones of the "cognitive synergy" idea, discussed repeatedly in prior chapters and more extensively in latter chapters, we now elaborate a little on the role it plays in the interaction between procedural and declarative learning.

While MOSES handles much of CogPrime's procedural learning, and CogPrime's internal simulation engine handles most episodic knowledge, CogPrime's primary tool for handling declarative knowledge is an uncertain inference framework called Probabilistic Logic Networks (PLN). The complexities of PLN are the topic of a lengthy technical monograph [GMIH08], and are summarized in Chapter 34; here we will eschew most details and focus mainly on pointing out how PLN seeks to achieve efficient inference control via integration with other cognitive processes.

As a logic, PLN is broadly integrative: it combines certain term logic rules with more standard predicate logic rules, and utilizes both fuzzy truth values and a variant of imprecise probabilities called *indefinite probabilities*. PLN mathematics tells how these uncertain truth values propagate through its logic rules, so that uncertain premises give rise to conclusions with reasonably accurately estimated uncertainty values. This careful management of uncertainty is critical for the application of logical inference in the robotics context, where most knowledge is abstracted from experience and is hence highly uncertain.

PLN can be used in either forward or backward chaining mode; and in the language introduced above, it can be used for either analysis or synthesis. As an example, we will consider backward chaining analysis, exemplified by the problem of a robot preschool-student trying to determine whether a new playmate "Bob" is likely to be a regular visitor to is preschool or not (evaluating the truth value of the implication $Bob \rightarrow regular\_visitor$). The basic backward chaining process for PLN analysis looks like:

1. Given an implication $L \equiv A \rightarrow B$ whose truth value must be estimated (for instance $L \equiv Concept \wedge Procedure \rightarrow Goal$ as discussed above), create a list $(A_1, ..., A_n)$ of *(inference rule, stored knowledge)* pairs that might be used to produce L
2. Using analogical reasoning to prior inferences, assign each A_i a probability of success
   - If some of the A_i are estimated to have reasonable probability of success at generating reasonably confident estimates of L's truth value, then invoke Step 1 with A_i in place of L (at this point the inference process becomes recursive)
   - If none of the A_i looks sufficiently likely to succeed, then inference has "gotten stuck" and another cognitive process should be invoked, e.g.
     - **Concept creation** may be used to infer new concepts related to A and B, and then Step 1 may be revisited, in the hope of finding a new, more promising A_i involving one of the new concepts
     - **MOSES** may be invoked with one of several special goals, e.g. the goal of finding a procedure P so that $P(X)$ predicts whether $X \rightarrow B$. If MOSES finds such a procedure P then this can be converted to declarative knowledge understandable by PLN and Step 1 may be revisited....
     - **Simulations** may be run in CogPrime's internal simulation engine, so as to observe the truth value of $A \rightarrow B$ in the simulations; and then Step 1 may be revisited....

The combinatorial explosion of inference control is combatted by the capability to defer to other cognitive processes when the inference control procedure is unable to make a sufficiently confident choice of which inference steps to take next. Note that just as MOSES may rely on PLN to model its evolving populations of procedures, PLN may rely on MOSES to create complex knowledge about the terms in its logical implications. This is just one example of the multiple ways in which the different cognitive processes in CogPrime interact synergetically; a more thorough treatment of these interactions is given in [Goe09a].

In the "new playmate" example, the interesting case is where the robot initially seems not to know enough about Bob to make a solid inferential judgment (so that none of the A_i seem particularly promising). For instance, it might carry out a number of possible inferences and not come to any reasonably confident conclusion, so that the reason none of the A_i seem promising is that all the decent-looking ones have been tried already. So it might then recourse to MOSES, simulation or concept creation.

For instance, the PLN controller could make a list of everyone who has been a regular visitor, and everyone who has not been, and pose MOSES the task of figuring out a procedure for distinguishing these two categories. This procedure could then be used directly to make the needed assessment, or else be translated into logical rules to be used within PLN inference. For example, perhaps MOSES would discover that older males wearing ties tend not to become regular visitors. If the new playmate is an older male wearing a tie, this is directly applicable. But if the current playmate is wearing a tuxedo, then PLN may be helpful via reasoning that even though a tuxedo is not a tie, it's a similar form of fancy dress – so PLN may extend the MOSES-learned rule to the present case and infer that the new playmate is not likely to be a regular visitor.

6.5 Goal-Oriented Dynamics in CogPrime

CogPrime's dynamics has both goal-oriented and "spontaneous" aspects; here for simplicity's sake we will focus on the goal-oriented ones. The basic goal-oriented dynamic of the CogPrime system, within which the various types of memory are utilized, is driven by implications known as "cognitive schematics", which take the form

$$Context \wedge Procedure \rightarrow Goal < p >$$

(summarized $C \wedge P \rightarrow G$). Semi-formally, this implication may be interpreted to mean: "If the context C appears to hold currently, then if I enact the procedure P, I can expect to achieve the goal G with certainty p." Cognitive synergy means that the learning processes corresponding to the different types of memory actively cooperate in figuring out what procedures will achieve the system's goals in the relevant contexts within its environment.

CogPrime's cognitive schematic is significantly similar to production rules in classical architectures like SOAR and ACT-R (as reviewed in Chapter 4; however, there are significant differences which are important to CogPrime's functionality. Unlike with classical production rules systems, uncertainty is core to CogPrime's knowledge representation, and each CogPrime cognitive schematic is labeled with an uncertain truth value, which is critical to its utilization by CogPrime's cognitive processes. Also, in CogPrime, cognitive schematics may be incomplete, missing one or two of the terms, which may then be filled in by various cognitive processes (generally in an uncertain way). A stronger similarity is to MicroPsi's triplets; the differences in this case are more low-level and technical and have already been mentioned in Chapter 4.

Finally, the biggest difference between CogPrime's cognitive schematics and production rules or other similar constructs, is that in CogPrime this level of knowledge representation is not the only important one. CLARION [SZ04], as reviewed above, is an example of a cognitive architecture that uses production rules for explicit knowledge representation and then uses a totally separate subsymbolic knowledge store for implicit knowledge. In CogPrime

both explicit and implicit knowledge are stored in the same graph of nodes and links, with

- explicit knowledge stored in probabilistic logic based nodes and links such as cognitive schematics (see Figure 6.8 for a depiction of some explicit linguistic knowledge.)
- implicit knowledge stored in patterns of activity among these same nodes and links, defined via the activity of the "importance" values (see Figure 6.9 for an illustrative example thereof) associated with nodes and links and propagated by the ECAN attention allocation process

The meaning of a cognitive schematic in CogPrime is hence not entirely encapsulated in its explicit logical form, but resides largely in the activity patterns that ECAN causes its activation or exploration to give rise to. And this fact is important because the synergetic interactions of system components are in large part modulated by ECAN activity. Without the real-time combination of explicit and implicit knowledge in the system's knowledge graph, the synergetic interaction of different cognitive processes would not work so smoothly, and the emergence of effective high-level hierarchical, heterarchical and self structures would be less likely.

6.6 Analysis and Synthesis Processes in CogPrime

We now return to CogPrime's fundamental cognitive dynamics, using examples from the "virtual dog" application to motivate the discussion.

The cognitive schematic $\mathit{Context} \wedge \mathit{Procedure} \rightarrow \mathit{Goal}$ leads to a conceptualization of the internal action of an intelligent system as involving two key categories of learning:

- **Analysis**: Estimating the probability p of a posited $C \wedge P \rightarrow G$ relationship
- **Synthesis**: Filling in one or two of the variables in the cognitive schematic, given assumptions regarding the remaining variables, and directed by the goal of maximizing the probability of the cognitive schematic

More specifically, where synthesis is concerned,

- The MOSES probabilistic evolutionary program learning algorithm is applied to find P, given fixed C and G. Internal simulation is also used, for the purpose of creating a simulation embodying C and seeing which P lead to the simulated achievement of G.
    - *Example: A virtual dog learns a procedure P to please its owner (the goal G) in the context C where there is a ball or stick present and the owner is saying "fetch".*
- PLN inference, acting on declarative knowledge, is used for choosing C, given fixed P and G (also incorporating sensory and episodic knowledge as appropriate). Simulation may also be used for this purpose.
    - *Example: A virtual dog wants to achieve the goal G of getting food, and it knows that the procedure P of begging has been successful at this before, so it seeks a context C where begging can be expected to get it food. Probably this will be a context involving a friendly person.*
- PLN-based goal refinement is used to create new subgoals G to sit on the right hand side of instances of the cognitive schematic.
    - *Example: Given that a virtual dog has a goal of finding food, it may learn a subgoal of following other dogs, due to observing that other dogs are often heading toward their food.*
- Concept formation heuristics are used for choosing G and for fueling goal refinement, but especially for choosing C (via providing new candidates for C). They are also used for choosing P, via a process called "predicate schematization" that turns logical predicates (declarative knowledge) into procedures.
    - *Example: At first a virtual dog may have a hard time predicting which other dogs are going to be mean to it. But it may eventually observe common features among a number of mean dogs, and thus form its own concept of "pit bull," without anyone ever teaching it this concept explicitly.*

Where analysis is concerned:

- PLN inference, acting on declarative knowledge, is used for estimating the probability of the implication in the cognitive schematic, given fixed C, P and G. Episodic knowledge is also used in this regard, via enabling estimation of the probability via simple similarity matching against past experience. Simulation is also used: multiple simulations may be run, and statistics may be captured therefrom.
  - *Example: To estimate the degree to which asking Bob for food (the procedure P is "asking for food", the context C is "being with Bob") will achieve the goal G of getting food, the virtual dog may study its memory to see what happened on previous occasions where it or other dogs asked Bob for food or other things, and then integrate the evidence from these occasions.*
- Procedural knowledge, mapped into declarative knowledge and then acted on by PLN inference, can be useful for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C \wedge P_1 \rightarrow G$ is known for some P_1 related to P.
  - *Example: knowledge of the internal similarity between the procedure of asking for food and the procedure of asking for toys, allows the virtual dog to reason that if asking Bob for toys has been successful, maybe asking Bob for food will be successful too.*
- Inference, acting on declarative or sensory knowledge, can be useful for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C_1 \wedge P \rightarrow G$ is known for some C_1 related to C.
  - *Example: if Bob and Jim have a lot of features in common, and Bob often responds positively when asked for food, then maybe Jim will too.*
- Inference can be used similarly for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C \wedge P \rightarrow G_1$ is known for some G_1 related to G. Concept creation can be useful indirectly in calculating these probability estimates, via providing new concepts that can be used to make useful inference trails more compact and hence easier to construct.
  - *Example: The dog may reason that because Jack likes to play, and Jack and Jill are both children, maybe Jill likes to play too. It can carry out this reasoning only if its concept creation process has invented the concept of "child" via analysis of observed data.*

In these examples we have focused on cases where two terms in the cognitive schematic are fixed and the third must be filled in; but just as often, the situation is that only one of the terms is fixed. For instance, if we fix G, sometimes the best approach will be to collectively learn C and P. This requires either a procedure learning method that works interactively with a declarative-knowledge-focused concept learning or reasoning method; or a declarative learning method that works interactively with a procedure learning method. That is, it requires the sort of cognitive synergy built into the CogPrime design.

6.7 Conclusion

To thoroughly describe a comprehensive, integrative AGI architecture in a brief chapter would be an impossible task; all we have attempted here is a brief overview, to be elaborated on in the 800-odd pages of Part 2 of this book. We do not expect this brief summary to be enough to convince the skeptical reader that the approach described here has a reasonable odds of success at achieving its stated goals, or even of fulfilling the conceptual notions outlined in the preceding chapters. However, we hope to have given the reader at least a rough idea of *what sort of AGI design* we are advocating, and *why and in what sense we believe it can lead to advanced artificial general intelligence.* For more details on the structure, dynamics and underlying concepts of CogPrime, the reader is encouraged to proceed to Part 2– after completing Part 1, of course. Please be patient – building a thinking machine is a big topic, and we have a lot to say about it!

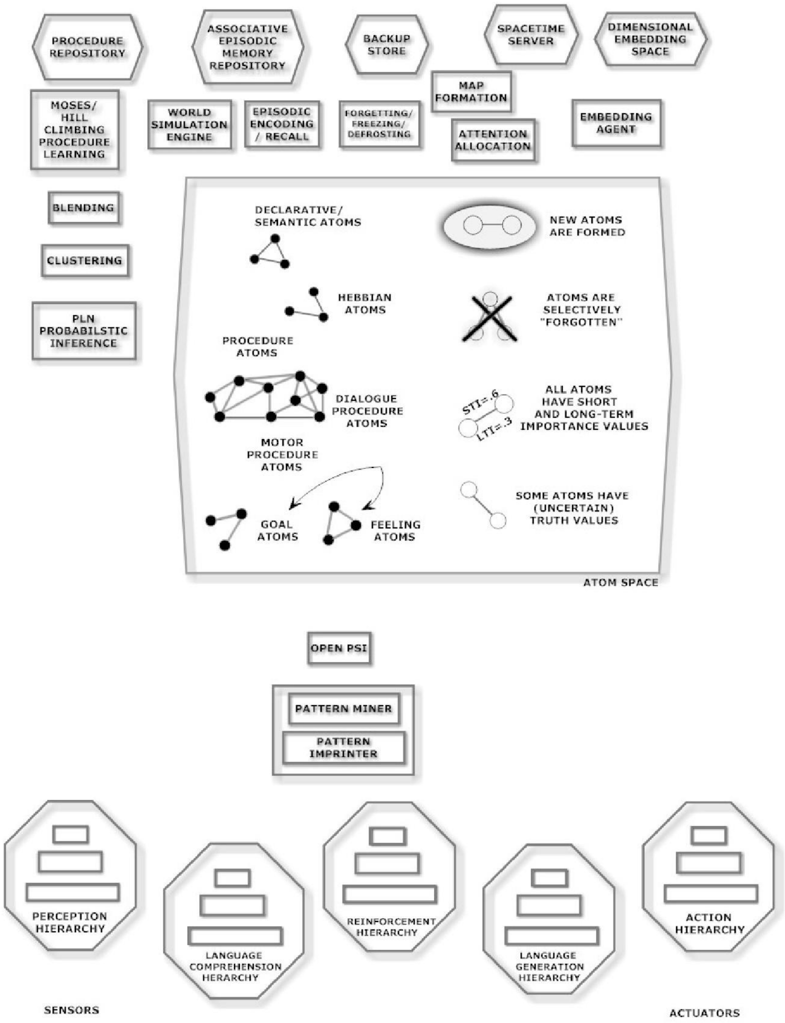

PROCEDURE REPOSITORY
ASSOCIATIVE EPISODIC MEMORY REPOSITORY
BACKUP STORE
SPACETIME SERVER
DIMENSIONAL EMBEDDING SPACE
MAP FORMATION
MOSES/ HILL CLIMBING PROCEDURE LEARNING
WORLD SIMULATION ENGINE
EPISODIC ENCODING / RECALL
FORGETTING/ FREEZING/ DEFROSTING
ATTENTION ALLOCATION
EMBEDDING AGENT
BLENDING
CLUSTERING
PLN PROBABILSTIC INFERENCE
DECLARATIVE/ SEMANTIC ATOMS
NEW ATOMS ARE FORMED
HEBBIAN ATOMS
ATOMS ARE SELECTIVELY "FORGOTTEN"
PROCEDURE ATOMS
DIALOGUE PROCEDURE ATOMS
STI=.6
LTI=.3
ALL ATOMS HAVE SHORT AND LONG-TERM IMPORTANCE VALUES
MOTOR PROCEDURE ATOMS
SOME ATOMS HAVE (UNCERTAIN) TRUTH VALUES
GOAL ATOMS
FEELING ATOMS
ATOM SPACE
OPEN PSI
PATTERN MINER
PATTERN IMPRINTER
PERCEPTION HIERARCHY
LANGUAGE COMPREHENSION HERARCHY
REINFORCEMENT HIERARCHY
LANGUAGE GENERATION HIERARCHY
ACTION HIERARCHY
SENSORS
ACTUATORS

Fig. 6.2: **Key Explicitly Implemented Processes of CogPrime** . The large box at the center is the Atomspace, the system's central store of various forms of (long-term and working) memory, which contains a weighted labeled hypergraph whose nodes and links are "Atoms" of various sorts. The hexagonal boxes at the bottom denote various hierarchies devoted to recognition and generation of patterns: perception, action and linguistic. Intervening between these recognition/generation hierarchies and the Atomspace, we have a pattern mining/imprinting component (that recognizes patterns in the hierarchies and passes them to the Atomspace; and imprints patterns from the Atomspace on the hierarchies); and also OpenPsi, a special dynamical framework for choosing actions based on motivations. Above the Atomspace we have a host of cognitive processes, which act on the Atomspace, some continually and some only as context dictates, carrying out various sorts of learning and reasoning (pertinent to various sorts of memory) that help the system fulfill its goals and motivations.

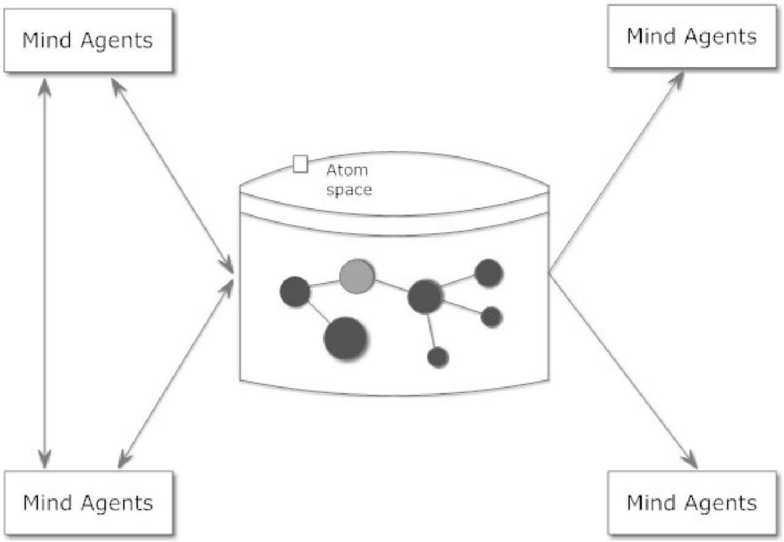

Mind Agents
Mind Agents
Atom space
Mind Agents
Mind Agents

Fig. 6.3: **MindAgents and AtomSpace in OpenCog.** This is a conceptual depiction of one way cognitive processes may interact in OpenCog – they may be wrapped in MindAgent objects, which interact via cooperatively acting on the AtomSpace.

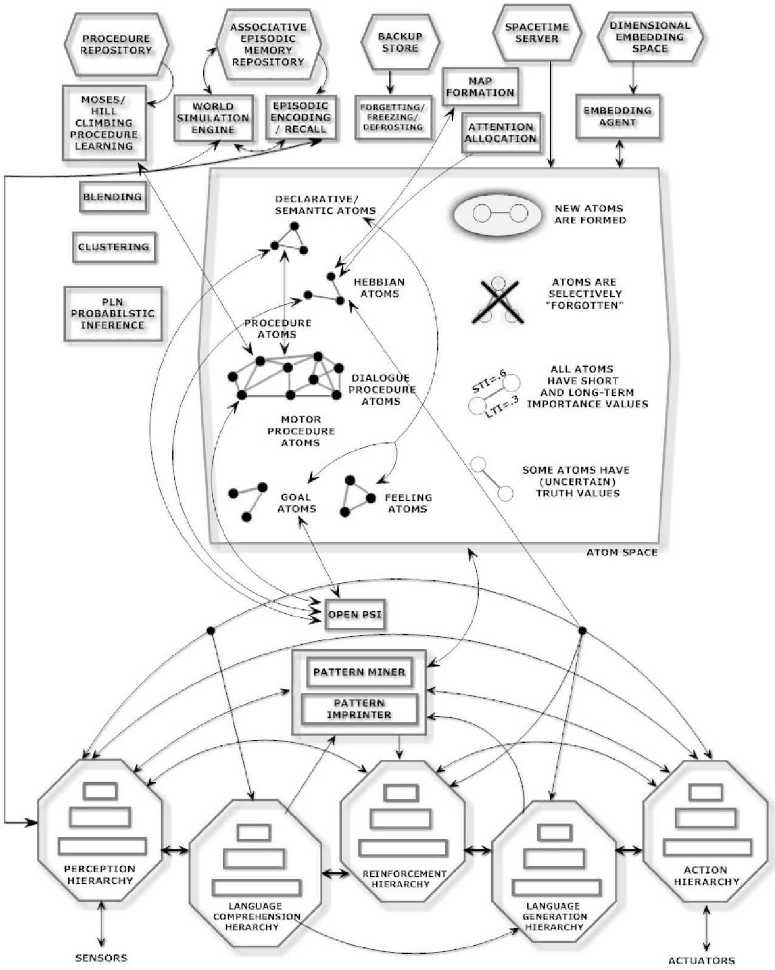

PROCEDURE REPOSITORY
ASSOCIATIVE EPISODIC MEMORY REPOSITORY
BACKUP STORE
SPACETIME SERVER
DIMENSIONAL EMBEDDING SPACE
MOSES/ HILL CLIMBING PROCEDURE LEARNING
WORLD SIMULATION ENGINE
EPISODIC ENCODING / RECALL
FORGETTING/ FREEZING/ DEFROSTING
MAP FORMATION
ATTENTION ALLOCATION
EMBEDDING AGENT
BLENDING
CLUSTERING
PLN PROBABILSTIC INFERENCE
DECLARATIVE/ SEMANTIC ATOMS
HEBBIAN ATOMS
PROCEDURE ATOMS
DIALOGUE PROCEDURE ATOMS
MOTOR PROCEDURE ATOMS
GOAL ATOMS
FEELING ATOMS
NEW ATOMS ARE FORMED
ATOMS ARE SELECTIVELY "FORGOTTEN"
STI=.6
LTI=.3
ALL ATOMS HAVE SHORT AND LONG-TERM IMPORTANCE VALUES
SOME ATOMS HAVE (UNCERTAIN) TRUTH VALUES
ATOM SPACE
OPEN PSI
PATTERN MINER
PATTERN IMPRINTER
PERCEPTION HIERARCHY
LANGUAGE COMPREHENSION HERARCHY
REINFORCEMENT HIERARCHY
LANGUAGE GENERATION HIERARCHY
ACTION HIERARCHY
SENSORS
ACTUATORS

Fig. 6.4: **Links Between Cognitive Processes and the Atomspace**. The cognitive processes depicted all act on the Atomspace, in the sense that they operate by observing certain Atoms in the Atomspace and then modifying (or in rare cases deleting) them, and potentially adding new Atoms as well. Atoms represent all forms of knowledge, but some forms of knowledge are additionally represented by external data stores connected to the Atomspace, such as the Procedure Repository; these are also shown as linked to the Atomspace.

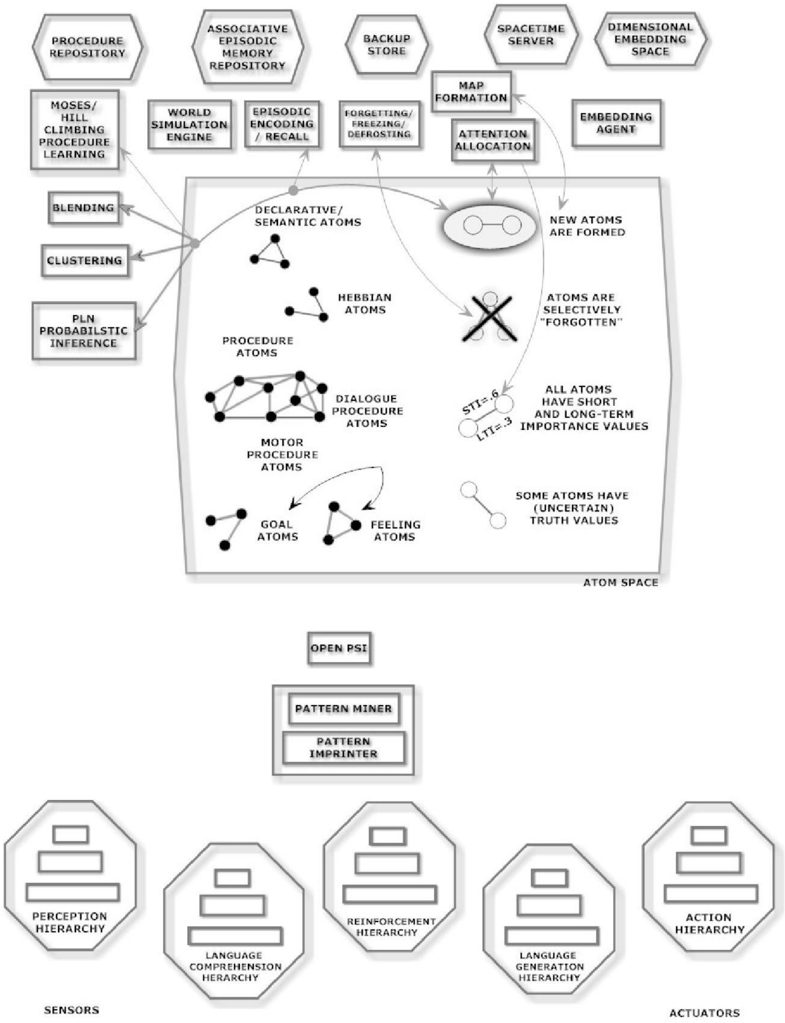
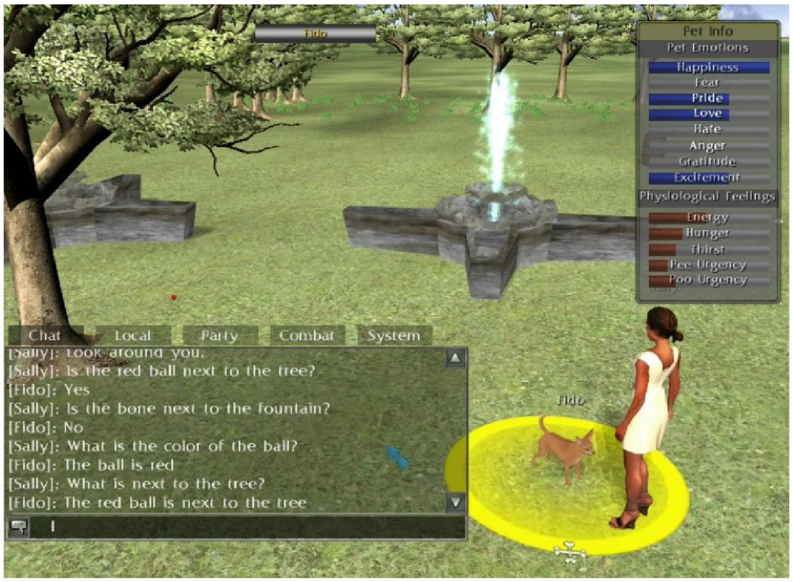

PROCEDURE REPOSITORY
ASSOCIATIVE EPISODIC MEMORY REPOSITORY
BACKUP STORE
SPACETIME SERVER
DIMENSIONAL EMBEDDING SPACE
MAP FORMATION
MOSES/ HILL CLIMBING PROCEDURE LEARNING
WORLD SIMULATION ENGINE
EPISODIC ENCODING / RECALL
FORGETTING/ FREEZING/ DEFROSTING
ATTENTION ALLOCATION
EMBEDDING AGENT
BLENDING
CLUSTERING
PLN PROBABILSTIC INFERENCE
DECLARATIVE/ SEMANTIC ATOMS
NEW ATOMS ARE FORMED
HEBBIAN ATOMS
ATOMS ARE SELECTIVELY "FORGOTTEN"
PROCEDURE ATOMS
DIALOGUE PROCEDURE ATOMS
STI=.6
LTI=.3
ALL ATOMS HAVE SHORT AND LONG-TERM IMPORTANCE VALUES
MOTOR PROCEDURE ATOMS
SOME ATOMS HAVE (UNCERTAIN) TRUTH VALUES
GOAL ATOMS
FEELING ATOMS
ATOM SPACE
OPEN PSI
PATTERN MINER
PATTERN IMPRINTER
PERCEPTION HIERARCHY
LANGUAGE COMPREHENSION HERARCHY
REINFORCEMENT HIERARCHY
LANGUAGE GENERATION HIERARCHY
ACTION HIERARCHY
SENSORS
ACTUATORS

Fig. 6.5: **Invocation of Atom Operations By Cognitive Processes**. This diagram depicts some of the Atom modification, creation and deletion operations carried out by the abstract cognitive processes in the CogPrime architecture.

CogPrime Component	**Int. Diag. Sub-Diagram**	**Int. Diag. Component**
Procedure Repository	Long-Term Memory	Procedural
Procedure Repository	Working Memory	Active Procedural
Associative Episodic Memory	Long-Term Memory	Episodic
Associative Episodic Memory	Working Memory	Transient Episodic
Backup Store	Long-Term Memory	*no correlate: a function not necessarily possessed by the human mind*
Spacetime Server	Long-Term Memory	Declarative and Sensorimotor
Dimensional Embedding Space	*no clear correlate: a tool for helping multiple types of LTM*	
Dimensional Embedding Agent	*no clear correlate*	
Blending	Long-Term and Working Memory	Concept Formation
Clustering	Long-Term and Working Memory	Concept Formation
PLN Probabilistic Inference	Long-Term and Working Memory	Reasoning and Plan Learning/Optimization
MOSES / Hillclimbing	Long-Term and Working Memory	Procedure Learning
World Simulation	Long-Term and Working Memory	Simulation
Episodic Encoding / Recall	Long-Term g Memory	Story-telling
Episodic Encoding / Recall	Working Memory	Consolidation
Forgetting / Freezing / Defrosting	Long-Term and Working Memory	*no correlate: a function not necessarily possessed by the human mind*
Map Formation	Long-Term Memory	Concept Formation and Pattern Mining
Attention Allocation	Long-Term and Working Memory	Hebbian/Attentional Learning
Attention Allocation	High-Level Mind Architecture	Reinforcement
Attention Allocation	Working Memory	Perceptual Associative Memory and Local Association
AtomSpace	High-Level Mind Architecture	*no clear correlate: a general tool for representing memory including long-term and working, plus some of perception and action*
AtomSpace	Working Memory	Global Workspace (the high-STI portion of AtomSpace) & other Workspaces
Declarative Atoms	Long-Term and Working Memory	Declarative and Sensorimotor
Procedure Atoms	Long-Term and Working Memory	Procedural
Hebbian Atoms	Long-Term and Working Memory	Attentional
Goal Atoms	Long-Term and Working Memory	Intentional
Feeling Atoms	Long-Term and Working Memory	spanning Declarative, Intentional and Sensorimotor
OpenPsi	High-Level Mind Architecture	Motivation / Action Selection
OpenPsi	Working Memory	Action Selection
Pattern Miner	High-Level Mind Architecture	arrows between perception and working and long-term memory
Pattern Miner	Working Memory	arrows between sensory memory and perceptual associative and transient episodic memory


Fido
Pet Info
Pet Emotions
Happiness
Fear
Pride
Love
Hate
Anger
Gratitude
Excitement
Physiological Feelings
Energy
Hunger
Thirst
Pee Urgency
Poo Urgency
Chat
Local
Party
Combat
System
Fido

[Sally]: Look around you.
[Sally]: Is the red ball next to the tree?
[Fido]: Yes
[Sally]: Is the bone next to the fountain?
[Fido]: No
[Sally]: What is the color of the ball?
[Fido]: The ball is red
[Sally]: What is next to the tree?
[Fido]: The red ball is next to the tree

Fig. 6.6: Screenshot of OpenCog-controlled virtual dog

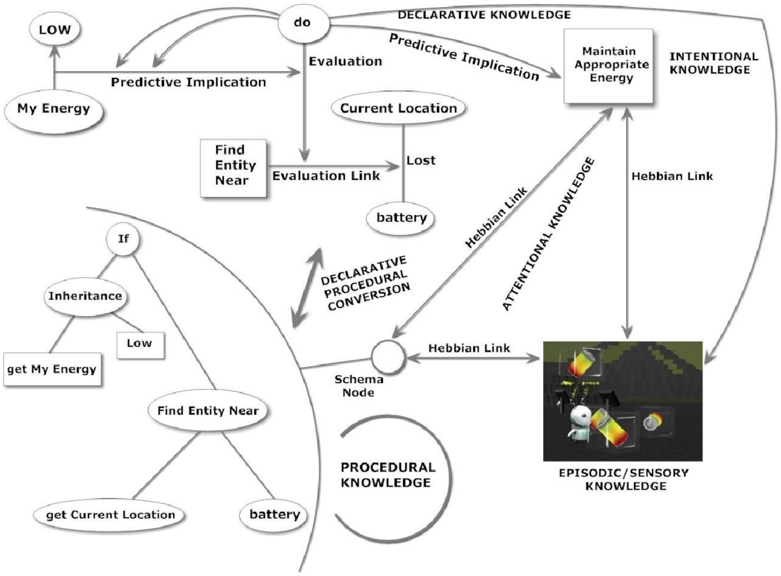

LOW
do
DECLARATIVE KNOWLEDGE
Predictive Implication
Evaluation
Predictive Implication
Maintain Appropriate Energy
INTENTIONAL KNOWLEDGE
My Energy
Current Location
Find Entity Near
Evaluation Link
Lost
battery
Hebbian Link
Hebbian Link
ATTENTIONAL KNOWLEDGE
If
Inheritance
DECLARATIVE PROCEDURAL CONVERSION
Low
get My Energy
Schema Node
Hebbian Link
Find Entity Near
PROCEDURAL KNOWLEDGE
EPISODIC/SENSORY KNOWLEDGE
get Current Location
battery

Fig. 6.7: **Relationship Between Multiple Memory Types**. The bottom left corner shows a program tree, constituting procedural knowledge. The upper left shows declarative nodes and links in the Atomspace. The upper right corner shows a relevant system goal. The lower right corner contains an image symbolizing relevant episodic and sensory knowledge. All the various types of knowledge link to each other and can be approximatively converted to each other.

Memory Type	**Specific Cognitive Processes**	**General Cognitive Functions**
Declarative	Probabilistic Logic Networks (PLN) [GMIH08]; conceptual blending [FT02]	pattern creation
Procedural	MOSES (a novel probabilistic evolutionary program learning algorithm) [Loo06]	pattern creation
Episodic	internal simulation engine [GEA08]	association, pattern creation
Attentional	Economic Attention Networks (ECAN) [GPI+10]	association, credit assignment
Intentional	probabilistic goal hierarchy refined by PLN and ECAN, structured according to MicroPsi [Bac09]	credit assignment, pattern creation
Sensory	In CogBot, this will be supplied by the DeSTIN component	association, attention allocation, pattern creation, credit assignment

Table 6.2: Memory Types and Cognitive Processes in CogPrime. The third column indicates the general cognitive function that each specific cognitive process carries out, according to the patternist theory of cognition.

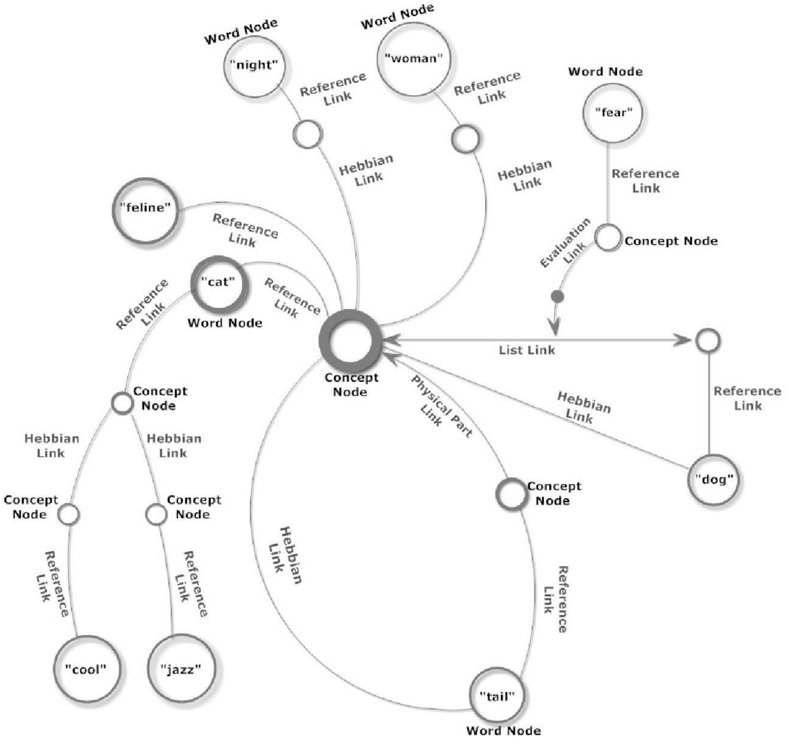

Word Node
"night"
Reference Link
Hebbian Link
Word Node
"woman"
Reference Link
Hebbian Link
Word Node
"fear"
Reference Link
Concept Node
Evaluation Link
"feline"
Reference Link
"cat"
Word Node
Reference Link
Reference Link
Concept Node
List Link
Reference Link
Hebbian Link
Physical Part Link
Concept Node
Hebbian Link
Hebbian Link
Concept Node
Concept Node
"dog"
Concept Node
Reference Link
Reference Link
Hebbian Link
Reference Link
"cool"
"jazz"
"tail"
Word Node

Fig. 6.8: **Example of Explicit Knowledge in the Atomspace**. One simple example of explicitly represented knowledge in the Atomspace is linguistic knowledge, such as words and the concepts directly linked to them. Not all of a CogPrime system's concepts correlate to words, but some do.

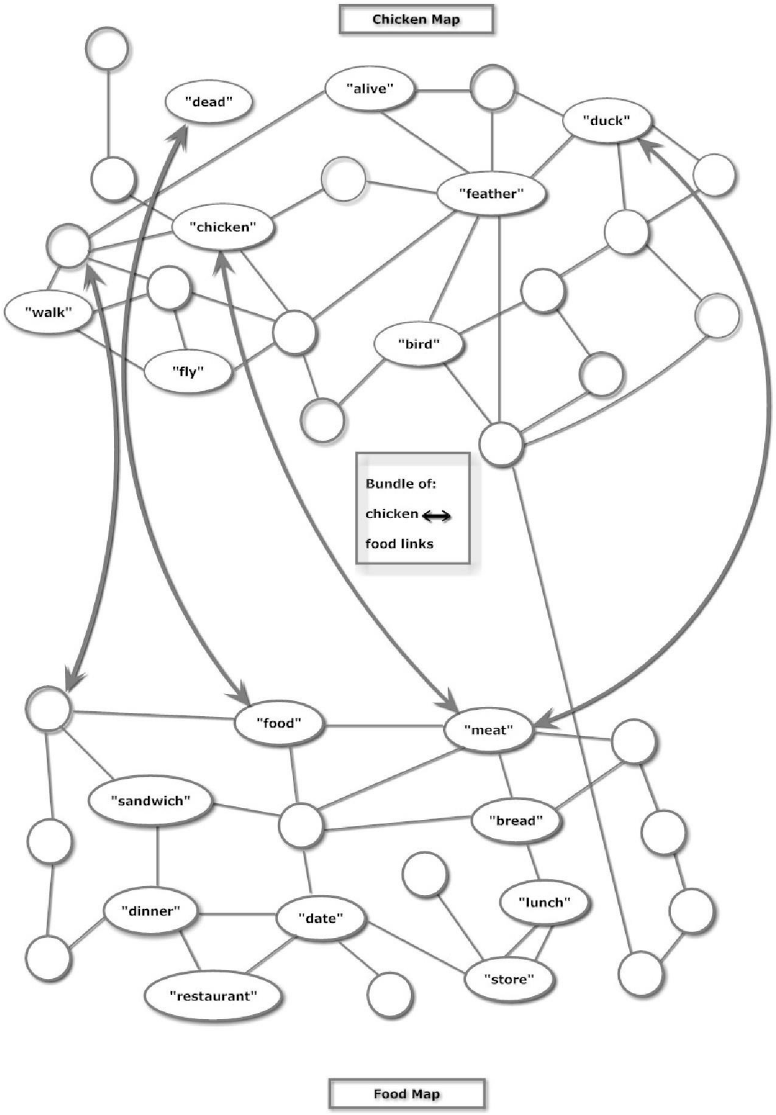

Chicken Map
"dead"
"alive"
"duck"
"feather"
"chicken"
"walk"
"fly"
"bird"
Bundle of:
chicken ⟷
food links
"food"
"meat"
"sandwich"
"bread"
"lunch"
"dinner"
"date"
"store"
"restaurant"
Food Map

Fig. 6.9: **Example of Implicit Knowledge in the Atomspace**. A simple example of implicit knowledge in the Atomspace. The "chicken" and "food" concepts are represented by "maps" of ConceptNodes interconnected by HebbianLinks, where the latter tend to form between ConceptNodes that are often simultaneously important. The bundle of links between nodes in the chicken map and nodes in the food map, represents an "implicit, emergent link" between the two concept maps. This diagram also illustrates "glocal" knowledge representation, in that the chicken and food concepts are each represented by individual nodes, but also by distributed maps. The "chicken" ConceptNode, when important, will tend to make the rest of the map important – and vice versa. Part of the overall chicken concept possessed by the system is expressed by the explicit links coming out of the chicken ConceptNode, and part is represented only by the distributed chicken map as a whole.

Section II

Toward a General Theory of General Intelligence

Chapter 7
A Formal Model of Intelligent Agents

7.1 Introduction

The artificial intelligence field is full of sophisticated mathematical models and equations, but most of these are highly specialized in nature – e.g. formalizations of particular logic systems, analyzes of the dynamics of specific sorts of neural nets, etc. On the other hand, a number of highly general models of intelligent systems also exist, including Hutter's recent formalization of universal intelligence [Hut05] and a large body of work in the disciplines of systems science and cybernetics – but these have tended not to yield many specific lessons useful for engineering AGI systems, serving more as conceptual models in mathematical form.

It would be fantastic to have a mathematical theory bridging these extremes – a real "general theory of general intelligence," allowing the derivation and analysis of specific structures and processes playing a role in practical AGI systems, from broad mathematical models of general intelligence in various situations and under various constraints. However, the path to such a theory is not entirely clear at present; and, as valuable as such a theory would be, we don't believe such a thing to be *necessary* for creating advanced AGI. One possibility is that the development of such a theory will occur contemporaneously and synergetically with the advent of practical AGI technology.

Lacking a mature, pragmatically useful "general theory of general intelligence," however, we have still found it valuable to articulate certain theoretical ideas about the nature of general intelligence, with a level of rigor a bit greater than the wholly informal discussions of the previous chapters. The chapters in this section of the book articulate some ideas we have developed in pursuit of a general theory of general intelligence; ideas that, even in their current relatively undeveloped form, have been very helpful in guiding our concrete work on the CogPrime design.

This chapter presents a more formal version of the notion of intelligence as "achieving complex goals in complex environments," based on a formal model of intelligent agents. These formalizations of agents and intelligence will be used in later chapters as a foundation for formalizing other concepts like inference and cognitive synergy. Chapters 8 and 9 pursue the notion of cognitive synergy a little more thoroughly than was done in previous chapters. Chapter 10 sketches a general theory of general intelligence using tools from category theory – not bringing it to the level where one can use it to derive specific AGI algorithms and structures; but still, presenting ideas that will be helpful in interpreting and explaining specific aspects of the CogPrime design in Part 2. Finally, Appendix **??** explores an additional theoretical direction, in which the mind of an intelligent system is viewed in terms of certain curved spaces – a novel way of thinking about the dynamics of general intelligence, which has been useful in guiding development of the ECAN component of CogPrime, and we expect will have more general value in future.

Despite the intermittent use of mathematical formalism, the ideas presented in this section are fairly speculative, and we do not propose them as constituting a well-demonstrated theory of general intelligence. Rather, we propose them as an interesting *way of thinking* about general intelligence, which appears to be consistent with available data, and which has proved inspirational to us in conceiving concrete structures and dynamics for AGI, as manifested for example in the CogPrime design. Understanding the way of thinking described in these chapters is valuable for understanding why the CogPrime design is the way it is, and for relating CogPrime to other practical and intellectual systems, and extending and improving CogPrime.

7.2 A Simple Formal Agents Model (SRAM)

We now present a formalization of the concept of "intelligent agents" – beginning with a formalization of "agents" in general.

Drawing on [Hut05, LH07a], we consider a class of active agents which observe and explore their environment and also take actions in it, which may affect the environment. Formally, the agent sends information to the environment by sending symbols from some finite alphabet called the *action space* Σ; and the environment sends signals to the agent with symbols from an alphabet called the *perception space*, denoted $\mathcal{P}$. Agents can also experience rewards, which lie in the *reward space*, denoted $\mathcal{R}$, which for each agent is a subset of the rational unit interval.

The agent and environment are understood to take turns sending signals back and forth, yielding a history of actions, observations and rewards, which may be denoted

$$a_1 o_1 r_1 a_2 o_2 r_2 ...$$

or else

$$a_1 x_1 a_2 x_2 ...$$

if x is introduced as a single symbol to denote both an observation and a reward. The complete interaction history up to and including cycle t is denoted $ax_{1:t}$; and the history before cycle t is denoted $ax_{<t} = ax_{1:t-1}$.

The agent is represented as a function π which takes the current history as input, and produces an action as output. Agents need not be deterministic, an agent may for instance induce a probability distribution over the space of possible actions, conditioned on the current history. In this case we may characterize the agent by a probability distribution $\pi(a_t | ax_{<t})$. Similarly, the environment may be characterized by a probability distribution $\mu(x_k | ax_{<k} a_k)$. Taken together, the distributions π and μ define a probability measure over the space of interaction sequences.

Next, we extend this model in a few ways, intended to make it better reflect the realities of intelligent computational agents. The first modification is to allow agents to maintain memories (of finite size), via adding memory actions drawn from a set $\mathcal{M}$ into the history of actions, observations and rewards. The second modification is to introduce the notion of **goals**.

7.2.1 Goals

We define goals as mathematical functions (to be specified below) associated with symbols drawn from the alphabet $\mathcal{G}$; and we consider the environment as sending goal-symbols to the agent along with regular observation-symbols. (Note however that the presentation of a goal-symbol to an agent does not necessarily entail the explicit communication to the agent of the contents of the goal function. This must be provided by other, correlated observations.) We also introduce a conditional distribution $\gamma(g, \mu)$ that gives the weight of a goal g in the context of a particular environment μ.

In this extended framework, an interaction sequence looks like

$$a_1 o_1 g_1 r_1 a_2 o_2 g_2 r_2 ...$$

or else

$$a_1 y_1 a_2 y_2 ...$$

where g_i are symbols corresponding to goals, and y is introduced as a single symbol to denote the combination of an observation, a reward and a goal.

Each goal function maps each finite interaction sequence $I_{g,s,t} = ay_{s:t}$ with g_s to g_t corresponding to g, into a value $r_g(I_{g,s,t}) \in [0, 1]$ indicating the value or "raw reward" of achieving the goal during that interaction sequence. The total reward r_t obtained by the agent is the sum of the raw rewards obtained at time t from all goals whose symbols occur in the agent's history before t.

This formalism of goal-seeking agents allows us to formalize the notion of intelligence as "achieving complex goals in complex environments" – a direction that is pursued in Section 7.3 below.

Note that this is an *external* perspective of system goals, which is natural from the perspective of formally defining system intelligence in terms of system behavior, but is not necessarily very natural in terms of system design. From the point of view of AGI design, one is generally more concerned with the (implicit or explicit) representation of goals inside an AGI system, as in CogPrime's Goal Atoms to be reviewed in Chapter 22 below.

Further, it is important to also consider the case where an AGI system has no explicit goals, and the system's environment has no immediately identifiable goals either. But in this case, we don't see any clear way to define a system's intelligence, except via *approximating* the system in terms of other theoretical systems which do have explicit goals. This approximation approach is developed in Section 7.3.5 below.

The awkwardness of linking the general formalism of intelligence theory presented here, with the practical business of creating and designing AGI systems, may indicate a shortcoming on the part of contemporary intelligence theory or AGI designs. On the other hand, this sort of situation often occurs in other domains as well – e.g. the leap from quantum theory to the analysis of real-world systems like organic molecules involves a lot of awkwardness and large leaps a well.

7.2.2 Memory Stores

As well as goals, we introduce into the model a long-term memory and a workspace. Regarding long-term memory we assume the agent's memory consists of multiple memory stores corresponding to various types of memory, e.g.: procedural (K_{Proc}), declarative (K_{Dec}), episodic (K_{Ep}), attentional (K_{Att}) and Intentional (K_{Int}). In Appendix **??** a category-theoretic model of these memory stores is introduced; but for the moment, we need only assume the existence of

- an injective mapping $\Theta_{Ep} : K_{Ep} \rightarrow \mathcal{H}$ where $\mathcal{H}$ is the space of fuzzy sets of subhistories (subhistories being "episodes" in this formalism)
- an injective mapping $\Theta_{Proc} : K_{Proc} \times \mathcal{M} \times \mathcal{W} \rightarrow \mathcal{A}$, where $\mathcal{M}$ is the set of memory states, $\mathcal{W}$ is the set of (observation, goal, reward) triples, and $\mathcal{A}$ is the set of actions (this maps each procedure object into a function that enacts actions in the environment or memory, based on the memory state and current world-state)
- an injective mapping $\Theta_{Dec} : K_{Dec} \rightarrow \mathcal{L}$, where $\mathcal{L}$ is the set of expressions in some formal language (which may for example be a logical language), which possesses words corresponding to the observations, goals, reward values and actions in our agent formalism
- an injective mapping $\Theta_{Int} : K_{Int} \rightarrow \mathcal{G}$, where $\mathcal{G}$ is the space of goals mentioned above
- an injective mapping $\Theta_{Att} : K_{Int} \cup K_{Ep} \cup K_{Proc} \cup K_{Ec} \rightarrow \mathcal{V}$, where $\mathcal{V}$ is the space of "attention values" (structures that gauge the importance of paying attention to an item of knowledge over various time-scales or in various contexts)

We also assume that the vocabulary of actions contains memory-actions corresponding to the operations of inserting the current observation, goal, reward or action into the episodic and/or declarative memory store. And, we assume that the activity of the agent, at each time-step, includes the enaction of one or more of the procedures in the procedural memory store. If several procedures are enacted at once, then the end result is still formally modeled as a single action $a = a_{[1]} * ... * a_{[k]}$ where $*$ is an operator on action-space that composes multiple actions into a single one.

Finally, we assume that, at each time-step, the agent may carry out an external action a_i on the environment, a memory action m_i on the (long-term) memory, and an action b_i on its **internal workspace**. Among the actions that can be carried out on the workspace, are the ability to insert or delete observations, goals, actions or reward-values from the workspace. The workspace can be thought of as a sort of short-term memory or else in terms of Baars' "global workspace" concept mentioned above. The workspace provides a medium for interaction between the different memory types.

The workspace provides a mechanism by which declarative, episodic and procedural memory may interact with each other. For this mechanism to work, we must assume that there are actions corresponding to query operations that allow procedures to look into declarative and episodic memory. The nature of these query operations will vary among different agents, but we can assume that in general an agent has

- one or more procedures $Q_{Dec}(x)$ serving as *declarative queries*, meaning that when Q_{Dec} is enacted on some x that is an ordered set of items in the workspace, the result is that one or more items from declarative memory is entered into the workspace
- one or more procedures $Q_{Ep}(x)$ serving as *episodic queries*, meaning that when Q_{Ep} is enacted on some x that is an ordered set of items in the workspace, the result is that one or more items from episodic memory is entered into the workspace

One additional aspect of CogPrime's knowledge representation that is important to PLN is the attachment of nonnegative weights n_i corresponding to elementary observations o_i. These weights denote the amount of evidence contained in the observation. For instance, in the context of a robotic agent, one could use these values to encode the assumption that an elementary visual observation has more evidential value than an elementary olfactory observation.

We now have a model of an agent with long-term memory comprising procedural, declarative and episodic aspects, an internal cognitive workspace, and the capability to use procedures to drive actions based on items in memory and the workspace, and to move items between long-term memory and the workspace.

7.2.2.1 Modeling CogPrime

Of course, this formal model may be realized differently in various real-world AGI systems. In CogPrime we have

- a weighted, labeled hypergraph structure called the AtomSpace used to store declarative knowledge (this is the representation used by PLN)
- a collection of programs in a LISP-like language called Combo, stored in a ProcedureRepository data structure, used to store procedural knowledge
- a collection of partial "movies" of the system's experience, played back using an internal simulation engine, used to store episodic knowledge
- AttentionValue objects, minimally containing ShortTermImportance (STI) and LongTermImportance (LTI) values used to store attentional knowledge
- Goal Atoms for intentional knowledge, stored in the same format as declarative knowledge but whose dynamics involve a special form of artificial currency that is used to govern action selection

The AtomSpace is the central repository and procedures and episodes are linked to Atoms in the AtomSpace which serve as their symbolic representatives. The "workspace" in CogPrime exists only virtually: each item in the AtomSpace has a "short term importance" (STI) level, and the workspace consists of those items in the AtomSpace with highest STI, and those procedures and episodes whose symbolic representatives in the AtomSpace have highest STI.

On the other hand, as we saw above, the LIDA architecture uses separate representations for procedural, declarative and episodic memory, but also has an explicit workspace component, where the most currently contextually relevant items from all different types of memory are gathered and used together in the course of actions. However, compared to CogPrime, it lacks comparably fine-grained methods for integrating the different types of memory.

Systematically mapping various existing cognitive architectures, or human brain structure, into this formal agents model would be a substantial though quite plausible exercise; but we will not undertake this here.

7.2.3 The Cognitive Schematic

Next we introduce an additional specialization into SRAM: the **cognitive schematic**, written informally as

$$Context \;\&\; Procedure \rightarrow Goal$$

and considered more formally as $holds(C)$ & $ex(P) \rightarrow h_i$ where h may be an externally specified goal g_i or an internally specified goal h derived as a (possibly uncertain) subgoal of one of more g_i; C is a piece of declarative or episodic knowledge and P is a procedure that the agent can internally execute to generate a series of actions. $ex(P)$ is the proposition that P is successfully executed. If C is episodic then $holds(C)$ may be interpreted as the current context (i.e. some finite slice of the agent's history) being similar to C; if C is declarative then $holds(C)$ may be interpreted as the truth value of C evaluated at the current context. Note that C may refer to some part of the world quite distant from the agent's current sensory observations; but it may still be formally evaluated based on the agent's history.

In the standard CogPrime notation as introduced formally in Chapter 20 (where indentation has function-argument syntax similar to that in Python, and relationship types are prepended to their relata without parentheses), for the case C is declarative this would be written as

```
PredictiveExtensionalImplication
   AND
       C
       Execution P
   G
```

and in the case C is episodic one replaces C in this formula with a predicate expressing C's similarity to the current context. The semantics of the PredictiveExtensionalInheritance relation will be discussed below. The Execution relation simply denotes the proposition that procedure P has been executed.

For the class of SRAM agents who (like CogPrime) use the cognitive schematic to govern many or all of their actions, a significant fragment of agent intelligence boils down to estimating the truth values of PredictiveExtensionalImplication relationships. Action selection procedures can be used, which choose procedures to enact based on which ones are judged most likely to achieve the current external goals g_i in the current context. Rather than enter into the particularities of action selection or other cognitive architecture issues, we will restrict ourselves to PLN inference, which in the context of the present agent model is a method for handling PredictiveImplication in the cognitive schematic.

Consider an agent in a virtual world, such as a virtual dog, one of whose external goals is to please its owner. Suppose its owner has asked it to find a cat, and it can translate this into a subgoal "find cat." If the agent operates according to the cognitive schematic, it will search for P so that

```
PredictiveExtensionalImplication
   AND
       C
       Execution P
   Evaluation
       found
       cat
```

holds.

7.3 Toward a Formal Characterization of Real-World General Intelligence

Having defined what we mean by an agent acting in an environment, we now turn to the question of what it means for such an agent to be "intelligent."

As we have reviewed extensively in Chapter 2 above, "intelligence" is a commonsense, "folk psychology" concept, with all the imprecision and contextuality that this generally entails. One cannot expect any compact, elegant formalism to capture all of its meanings. Even in the psychology and AI research communities, divergent definitions abound; Legg and Hutter [LH07a] lists and organizes 70+ definitions from the literature.

Practical study of natural intelligence in humans and other organisms, and practical design, creation and instruction of artificial intelligences, can proceed perfectly well without an agreed-upon formalization of the "intelligence" concept. Some researchers may conceive their own formalisms to guide their own work, others may feel no need for any such thing.

But nevertheless, it is of interest to seek formalizations of the concept of intelligence, which capture useful fragments of the commonsense notion of intelligence, and provide guidance for practical research in cognitive science and AI. A number of such formalizations have been given in recent decades, with varying degrees of mathematical rigor. Perhaps the most carefully-wrought formalization of intelligence so far is the theory of "universal intelligence" presented by Shane Legg and Marcus Hutter in [LH07b], which draws on ideas from algorithmic information theory.

Universal intelligence captures a certain aspect of the "intelligence" concept very well, and has the advantage of connecting closely with ideas in learning theory, decision theory and computation theory. However, the kind of general intelligence it captures best, is a kind which is in a sense *more general* in scope than human-style general intelligence. Universal intelligence does capture the sense in which humans are more intelligent than worms, which are more intelligent than rocks; and the sense in which theoretical AGI systems like Hutter's AIXI or $AIXI^{tl}$ [Hut05] would be much more intelligent than humans. But it misses essential aspects of the intelligence concept as it is used in the context of intelligent natural systems like humans or real-world AI systems.

Our main goal in this section is to present variants of universal intelligence that better capture the notion of intelligence as it is typically understood in the context of real-world natural and artificial systems. The first variant we describe is *pragmatic general intelligence*, which is inspired by the intuitive notion of intelligence as "the ability to achieve complex goals in complex environments," given in [Goe93a]. After assuming a prior distribution over the space of possible environments, and one over the space of possible goals, one then defines the pragmatic general intelligence as the expected level of goal-achievement of a system relative to these distributions. Rather than measuring truly broad mathematical general intelligence, pragmatic general intelligence measures intelligence in a way that's specifically biased toward certain environments and goals.

Another variant definition is then presented, the *efficient pragmatic general intelligence*, which takes into account the amount of computational resources utilized by the system in achieving its intelligence. Some argue that making efficient use of available resources is a defining characteristic of intelligence, see e.g. [Wan06].

A critical question left open is the characterization of the prior distributions corresponding to everyday human reality; we give a semi-formal sketch of some ideas on this in Chapter 9 below, where we present the notion of a "communication prior," which assigns a probability weight to a situation S based on the ease with which one agent in a society can communicate S to another agent in that society, using multimodal communication (including verbalization, demonstration, dramatic and pictorial depiction, etc.).

Finally, we present a formal measure of the "generality" of an intelligence, which precisiates the informal distinction between "general AI" and "narrow AI."

7.3.1 Biased Universal Intelligence

To define universal intelligence, Legg and Hutter consider the class of environments that are *reward-summable*, meaning that the total amount of reward they return to any agent is bounded by 1. Where r_i denotes the reward experienced by the agent from the environment at time i, the *expected total reward* for the agent π from the environment μ is defined as

$$V_\mu^\pi \equiv E(\sum_1^\infty r_i) \leq 1$$

To extend their definition in the direction of greater realism, we first introduce a second-order probability distribution ν, which is a probability distribution over the space of environments μ. The distribution ν assigns each environment a probability. One such distribution ν is the Solomonoff-Levin universal distribution in which one sets $\nu = 2^{-K(\mu)}$; but this is not the only distribution ν of interest. In fact a great deal of real-world general intelligence consists of the adaptation of intelligent systems to particular distributions ν over environment-space, differing from the universal distribution.

We then define

Definition 4 *The* **biased universal intelligence** *of an agent π is its expected performance with respect to the distribution ν over the space of all computable reward-summable environments, E, that is,*

$$\Upsilon(\pi) \equiv \sum_{\mu \in E} \nu(\mu) V_\mu^\pi$$

Legg and Hutter's **universal intelligence** is obtained by setting ν equal to the universal distribution.

This framework is more flexible than it might seem. E.g. suppose one wants to incorporate agents that die. Then one may create a special action, say a_{666}, corresponding to the state of death, to create agents that

- in certain circumstances output action a_{666}
- have the property that if their previous action was a_{666}, then all of their subsequent actions must be a_{666}

and to define a reward structure so that actions a_{666} always bring zero reward. It then follows that death is generally a bad thing if one wants to maximize intelligence. Agents that die will not get rewarded after they're dead; and agents that live only 70 years, say, will be restricted from getting rewards involving long-term patterns and will hence have specific limits on their intelligence.

7.3.2 Connecting Legg and Hutter's Model of Intelligent Agents to the Real World

A notable aspect of the Legg and Hutter formalism is the separation of the reward mechanism from the cognitive mechanisms of the agent. While commonplace in the reinforcement learning literature, this seems psychologically unrealistic in the context of biological intelligences and many types of machine intelligences. Not all human intelligent activity is specifically reward-seeking in nature; and even when it is, humans often pursue complexly constructed rewards, that are defined in terms of their own cognitions rather than separately given. Suppose a certain human's goals are true love, or world peace, and the proving of interesting theorems – then these goals are defined by the human herself, and only she knows if she's achieved them. An externally-provided reward signal doesn't capture the nature of this kind of goal-seeking behavior, which characterizes much human goal-seeking activity (and will presumably characterize much of the goal-seeking activity of advanced engineered intelligences also) ... let alone human behavior that is spontaneous and unrelated to explicit goals, yet may still appear commonsensically intelligent.

One could seek to bypass this complaint about the reward mechanisms via a sort of "neo-Freudian" argument, via

- associating the reward signal, not with the "external environment" as typically conceived, but rather with a portion of the intelligent agent's brain that is separate from the cognitive component
- viewing complex goals like true love, world peace and proving interesting theorems as indirect ways of achieving the agent's "basic goals", created within the agent's memory via subgoaling mechanisms

but it seems to us that a general formalization of intelligence should not rely on such strong assumptions about agents' cognitive architectures. So below, after introducing the pragmatic and efficient pragmatic general intelligence measures, we will propose an alternate interpretation wherein the mechanism of external rewards is viewed as a theoretical test framework for assessing agent intelligence, rather than a hypothesis about intelligent agent architecture.

In this alternate interpretation, formal measures like the universal, pragmatic and efficient pragmatic general intelligence are viewed as *not* directly applicable to real-world intelligences, because they involve the behaviors of agents over a wide variety of goals and environments, whereas in real life the opportunities to observe agents are more limited. However, they are viewed as being *indirectly* applicable to real-world agents, in the sense that an external intelligence can observe an agent's real-world behavior and then *infer* its likely intelligence according to these measures.

In a sense, this interpretation makes our formalized measures of intelligence the opposite of real-world IQ tests. An IQ test is a quantified, formalized test which is designed to approximately predict the informal, qualitative achievement of humans in real life. On the other hand, the formal definitions of intelligence we present here are quantified, formalized tests that are designed to capture abstract notions of intelligence, but which can be approximately evaluated on a real-world intelligent system by observing what it does in real life.

7.3.3 Pragmatic General Intelligence

The above concept of biased universal intelligence is perfectly adequate for many purposes, but it is also interesting to explicitly introduce the notion of a *goal* into the calculation. This allows us to formally capture the notion presented in [Goe93a] of intelligence as "the ability to achieve complex goals in complex environments."

If the agent is acting in environment μ, and is provided with g_s corresponding to g at the start and the end of the time-interval $T = \{i \in (s, ..., t)\}$, then the *expected goal-achievement* of the agent, relative to g, during the interval is the expectation

$$V^{\pi}_{\mu,g,T} \equiv E(\sum_{i=s}^{t} r_g(I_{g,s,i}))$$

where the expectation is taken over all interaction sequences $I_{g,s,i}$ drawn according to μ. We then propose

Definition 5 *The* **pragmatic general intelligence** *of an agent π, relative to the distribution ν over environments and the distribution γ over goals, is its expected performance with respect to goals drawn from γ in environments drawn from ν, over the time-scales natural to the goals; that is,*

$$\Pi(\pi) \equiv \sum_{\mu \in E, g \in \mathcal{G}, T} \nu(\mu)\gamma(g,\mu)V^{\pi}_{\mu,g,T}$$

(in those cases where this sum is convergent).

This definition formally captures the notion that "intelligence is achieving complex goals in complex environments," where "complexity" is gauged by the assumed measures ν and γ.

If ν is taken to be the universal distribution, and γ is defined to weight goals according to the universal distribution, then pragmatic general intelligence reduces to universal intelligence.

Furthermore, it is clear that a universal algorithmic agent like AIXI [Hut05] would also have a high pragmatic general intelligence, under fairly broad conditions. As the interaction history grows longer, the pragmatic general intelligence of AIXI would approach the theoretical maximum; as AIXI would implicitly infer the relevant distributions via experience. However, if significant reward discounting is involved, so that near-term rewards are weighted much higher than long-term rewards, then AIXI might compare very unfavorably in pragmatic general intelligence, to other agents designed with prior knowledge of ν, γ and τ in mind.

The most interesting case to consider is where ν and γ are taken to embody some particular bias in a real-world space of environments and goals, and this bias is appropriately reflected in the internal structure of an intelligent agent. Note that an agent needs not lack universal intelligence in order to possess pragmatic general intelligence with respect to some non-universal distribution over goals and environments. However, in general, given limited resources, there may be a tradeoff between universal intelligence and pragmatic intelligence. Which leads to the next point: how to encompass resource limitations into the definition.

One might argue that the definition of Pragmatic General Intelligence is already encompassed by Legg and Hutter's definition because one may bias the distribution of environments within the latter by considering different Turing machines underlying the Kolmogorov complexity. However this is not a general equivalence because the Solomonoff-Levin measure intrinsically decays exponentially, whereas an assumptive distribution over environments might decay at some other rate. This issue seems to merit further mathematical investigation.

7.3.4 Incorporating Computational Cost

Let $\eta_{\pi,\mu,g,T}$ be a probability distribution describing the amount of computational resources consumed by an agent π while achieving goal g over time-scale T. This is a probability distribution because we want to account for the possibility of nondeterministic agents. So, $\eta_{\pi,\mu,g,T}(Q)$ tells the probability that Q units of resources are consumed. For simplicity we amalgamate space and time resources, energetic resources, etc. into a single number Q, which is assumed to live in some subset of the positive reals. Space resources of course have to do with the size of the system's memory. Then we may define

Definition 6 *The* **efficient pragmatic general intelligence** *of an agent π with resource consumption $\eta_{\pi,\mu,g,T}$, relative to the distribution ν over environments and the distribution γ over goals, is its expected performance with respect to goals drawn from γ in environments drawn from ν, over the time-scales natural to the goals, normalized by the amount of computational effort expended to achieve each goal; that is,*

$$\Pi_{\mathit{Eff}}(\pi) \equiv \sum_{\mu\in E, g\in \mathcal{G}, Q, T} \frac{\nu(\mu)\gamma(g,\mu)\eta_{\pi,\mu,g,T}(Q)}{Q} V^{\pi}_{\mu,g,T}$$

(in those cases where this sum is convergent).

This is a measure that rates an agent's intelligence higher if it uses fewer computational resources to do its business. Roughly, it measures reward achieved per spacetime computation unit.

Note that, by abandoning the universal prior, we have also abandoned the proof of convergence that comes with it. In general the sums in the above definitions need not converge; and exploration of the conditions under which they do converge is a complex matter.

7.3.5 Assessing the Intelligence of Real-World Agents

The pragmatic and efficient pragmatic general intelligence measures are more "realistic" than the Legg and Hutter universal intelligence measure, in that they take into account the innate biasing and computational resource restrictions that characterize real-world intelligence. But as discussed earlier, they still live in "fantasy-land" to an extent – they gauge the intelligence of an agent via a weighted average over a wide variety of goals and environments; and they presume a simplistic relationship between agents and rewards that does not reflect the complexities of real-world cognitive architectures. It is not obvious from the foregoing how to apply these measures to real-world intelligent systems, which lack the ability to exist in such a wide variety of environments within their often brief lifespans, and mostly go about their lives doing things other than pursuing quantified external rewards. In this brief section we describe an approach to bridging this gap. The treatment is left semi-formal in places.

We suggest to view the definitions of pragmatic and efficient pragmatic general intelligence in terms of a "possible worlds" semantics – i.e. to view them as asking, counterfactually, how an agent *would* perform, hypothetically, on a series of tests (the tests being goals, defined in relation to environments and reward signals).

Real-world intelligent agents don't normally operate in terms of explicit goals and rewards; these are abstractions that we use to think about intelligent agents. However, this is no objection to characterizing various sorts of intelligence in terms of counterfactuals like: how would system S operate if it were trying to achieve this or that goal, in this or that environment, in order to seek reward? We can characterize various sorts of intelligence in terms of how it can be inferred an agent would perform on certain tests, even though the agent's real life does not consist of taking these tests.

This conceptual approach may seem a bit artificial but we don't currently see a better alternative, if one wishes to quantitatively gauge intelligence (which is, in a sense, an "artificial" thing to do in the first place). Given a real-world agent X and a mandate to assess its intelligence, the obvious alternative to looking at possible worlds in the manner of the above definitions, is just looking *directly* at the properties of the things X has achieved in the real world during its lifespan. But this isn't an easy solution, because it doesn't disambiguate which aspects of X's achievements were due to its own actions versus due to the rest of the world that X was interacting with when it made its achievements. To distinguish the amount of achievement that X "caused" via its own actions requires a model of causality, which is a complex can of worms in itself; and, critically, the standard models of causality also involve counterfactuals (asking "what would have been achieved in this situation if the agent X hadn't been there", etc.) [MW07]. Regardless of the particulars, it seems impossible to avoid counterfactual realities in assessing intelligence.

The approach we suggest – given a real-world agent X with a history of actions in a particular world, and a mandate to assess its intelligence – is to introduce an additional player, an *inference agent* δ, into the picture. The agent π modeled above is then viewed as π_X: the model of X that δ constructs, in order to explore X's inferred behaviors in various counterfactual environments. In the test situations embodied in the definitions of pragmatic and efficient pragmatic general intelligence, the environment gives π_X rewards, based on specifically configured goals. In X's real life, the relation between goals, rewards and actions will generally be significantly subtler and perhaps quite different.

We model the real world similarly to the "fantasy world" of the previous section, but with the omission of goals and rewards. We define a *naturalistic* context as one in which all goals and rewards are constant, i.e. $g_i = g_0$ and $r_i = r_0$ for all i. This is just a mathematical convention for stating that there are no precisely-defined external goals and rewards for the agent. In a naturalistic context, we then have a situation where agents create actions based on the past history of actions and perceptions, and if there is any relevant notion of reward or goal, it is within the cognitive mechanism of some agent. A *naturalistic agent* X is then an agent π which is restricted to one particular naturalistic context, involving one particular environment μ (formally, we may achieve this within the framework of agents described above via dictating that X issues constant "null actions" a_0 in all environments except μ).

Next, we posit a metric space (Σ_μ, d) of naturalistic agents defined on a naturalistic context involving environment μ, and a subspace $\Delta \in \Sigma_\mu$ of inference agents, which are naturalistic agents that output predictions of other agents' behaviors (a notion we will not fully formalize here). If agents are represented as program trees, then d may be taken as edit distance on tree space [Bil05]. Then, for each agent $\delta \in \Delta$, we may assess

- the prior probability $\theta(\delta)$ according to some assumed distribution θ
- the effectiveness $p(\delta, X)$ of δ at predicting the actions of an agent $X \in \Sigma_\mu$

We may then define

Definition 7 *The* **inference ability** *of the agent δ, relative to μ and X, is*

$$q_{\mu,X}(\delta) = \theta(\delta) \frac{\sum_{Y \in \Sigma_\mu} sim(X,Y) p(\delta, Y)}{\sum_{Y \in \Sigma_\mu} sim(X,Y)}$$

where sim is a specified decreasing function of $d(X,Y)$, such as $sim(X,Y) = \frac{1}{1+d(X,Y)}$.

To construct π_X, we may then use the model of X created by the agent $\delta \in \Delta$ with the highest inference ability relative to μ and X (using some specified ordering, in case of a tie). Having constructed π_X, we can then say that

Definition 8 *The inferred pragmatic general intelligence (relative to ν and γ) of a naturalistic agent X defined relative to an environment μ, is defined as the pragmatic general intelligence of the model π_X of X produced by the agent $\delta \in \Delta$ with maximal inference ability relative to μ (and in the case of a tie, the first of these in the ordering defined over Δ). The inferred efficient pragmatic general intelligence of X relative to μ is defined similarly.*

This provides a precise characterization of the pragmatic and efficient pragmatic intelligence of real-world systems, based on their observed behaviors. It's a bit messy; but the real world tends to be like that.

7.4 Intellectual Breadth: Quantifying the Generality of an Agent's Intelligence

We turn now to a related question: How can one quantify the degree of **generality** that an intelligent agent possesses? Above we have discussed the qualitative distinction between AGI and "Narrow AI", and intelligence as we have formalized it above is specifically intended as a measure of general intelligence. But quantifying intelligence is different than quantifying generality versus narrowness.

To make the discussion simpler, we introduce the term "context" as a shorthand for "environment/interval triple (μ, g, T)." Given a context (μ, g, T), and a set Σ of agents, one may construct a fuzzy set $Ag_{\mu,g,T}$ gathering those agents that are intelligent relative to the context; and given a set of contexts, one may also define a fuzzy set Con_π gathering those contexts with respect to which a given agent π is intelligent. The relevant formulas are:

$$\chi_{Ag_{\mu,g,T}}(\pi) = \chi_{Con_\pi}(\mu, g, T) = \frac{1}{N} \sum_Q \frac{\eta_{\mu,g,T}(Q) V^\pi_{\mu,g,T}}{Q}$$

where $N = N(\mu, g, T)$ is a normalization factor defined appropriately, e.g. via $N(\mu, g, T) = \max_\pi V^\pi_{\mu,g,T}$.

One could make similar definitions leaving out the computational cost factor Q, but we suspect that incorporating Q is a more promising direction. We then propose

Definition 9 *The* **intellectual breadth** *of an agent* π*, relative to the distribution* ν *over environments and the distribution* γ *over goals, is*

$$H(\chi^P_{Con_\pi}(\mu, g, T))$$

where H *is the entropy and*

$$\chi^P_{Con_\pi}(\mu, g, T) = \frac{\nu(\mu)\gamma(g, \mu)\chi_{Con_\pi}(\mu, g, T)}{\sum\limits_{(\mu_\alpha, g_\beta . T_\omega)} \nu(\mu_\alpha)\gamma(g_\beta, \mu_\alpha)\chi_{Con_\pi}(\mu_\alpha, g_\beta, T_\omega)}$$

is the probability distribution formed by normalizing the fuzzy set $\chi_{Con_\pi}(\mu, g, T)$*.*

A similar definition of the intellectual breadth of a context (μ, g, T), relative to the distribution σ over agents, may be posited. A weakness of these definitions is that they don't try to account for dependencies between agents or contexts; perhaps more refined formulations may be developed that account explicitly for these dependencies.

Note that the intellectual breadth of an agent as defined here is largely independent of the (efficient or not) pragmatic general intelligence of that agent. One could have a rather (efficiently or not) pragmatically generally intelligent system with little breadth: this would be a system very good at solving a fair number of hard problems, yet wholly incompetent on a larger number of hard problems. On the other hand, one could also have a terribly (efficiently or not) pragmatically generally stupid system with great intellectual breadth: i.e a system roughly equally dumb in all contexts!

Thus, one can characterize an intelligent agent as "narrow" with respect to distribution ν over environments and the distribution γ over goals, based on evaluating it as having low intellectual breadth. A "narrow AI" relative to ν and γ would then be an AI agent with a relatively high efficient pragmatic general intelligence but a relatively low intellectual breadth.

7.5 Conclusion

Our main goal in this chapter has been to push the formal understanding of intelligence in a more pragmatic direction. Much more work remains to be done, e.g. in specifying the environment, goal and efficiency distributions relevant to real-world systems, but we believe that the ideas presented here constitute nontrivial progress.

If the line of research suggested in this chapter succeeds, then eventually, one will be able to do AGI research as follows: Specify an AGI architecture formally, and then use the mathematics of general intelligence to derive interesting results about the environments, goals and hardware platforms relative to which the AGI architecture will display significant pragmatic or efficient pragmatic general intelligence, and intellectual breadth. The remaining chapters in this section present further ideas regarding how to work toward this goal. For the time being, such a mode of AGI research remains mainly for the future, but we have still found the formalism given in these chapters useful for formulating and clarifying various aspects of the CogPrime design as will be presented in later chapters.

Chapter 8
Cognitive Synergy

8.1 Cognitive Synergy

As we have seen, the formal theory of general intelligence, in its current form, doesn't really tell us much that's of use for creating real-world AGI systems. It tells us that creating extraordinarily powerful general intelligence is almost trivial if one has unrealistically huge amounts of computational resources; and that creating moderately powerful general intelligence using feasible computational resources is all about creating AI algorithms and data structures that (explicitly or implicitly) match the restrictions implied by a certain class of situations, to which the general intelligence is biased.

We've also described, in various previous chapters, some non-rigorous, conceptual principles that seem to explain key aspects of feasible general intelligence: the complementary reliance on evolution and autopoiesis, the superposition of hierarchical and heterarchical structures, and so forth. These principles can be considered as broad strategies for achieving general intelligence in certain broad classes of situations. Although, a lot of research needs to be done to figure out nice ways to describe, for instance, in what class of situations evolution is an effective learning strategy, in what class of situations dual hierarchical/heterarchical structure is an effective way to organize memory, etc.

In this chapter we'll dig deeper into one of the "general principle of feasible general intelligences" briefly alluded to earlier: the *cognitive synergy* principle, which is both a conceptual hypothesis about the structure of generally intelligent systems in certain classes of environments, and a design principle used to guide the architecting of CogPrime.

We will focus here on cognitive synergy specifically in the case of "multi-memory systems," which we define as intelligent systems (like CogPrime) whose combination of environment, embodiment and motivational systems make it important for them to possess memories that divide into partially but not wholly distinct components corresponding to the categories of:

- Declarative memory
- Procedural memory (memory about how to do certain things)
- Sensory and episodic memory
- Attentional memory (knowledge about what to pay attention to in what contexts
- Intentional memory (knowledge about the system's own goals and subgoals)

In Chapter 9 below we present a detailed argument as to how the requirement for a multi-memory underpinning for general intelligence emerges from certain underlying assumptions regarding the measurement of the simplicity of goals and environments; but the points made here do not rely on that argument. What they do rely on is the assumption that, in the intelligence in question, the different components of memory are significantly but not wholly distinct. That is, there are significant "family resemblances" between the memories of a single type, yet there are also thoroughgoing connections between memories of different types.

The cognitive synergy principle, if correct, applies to any AI system demonstrating intelligence in the context of embodied, social communication. However, one may also take the theory as an explicit guide for constructing AGI systems; and of course, the bulk of this book describes one AGI architecture, CogPrime, designed in such a way.

It is possible to cast these notions in mathematical form, and we make some efforts in this direction in Appendix **??**, using the languages of category theory and information geometry. However, this formalization has not yet led to any rigorous proof of the generality of cognitive synergy nor any other exciting theorems; with luck this will come as the mathematics is further developed. In this chapter the presentation is kept on the heuristic level, which is all that is critically needed for motivating the CogPrime design.

8.2 Cognitive Synergy

The essential idea of cognitive synergy, in the context of multi-memory systems, may be expressed in terms of the following points:

1. Intelligence, relative to a certain set of environments, may be understood as the capability to achieve complex goals in these environments.
2. With respect to certain classes of goals and environments (see Chapter 9 for a hypothesis in this regard), an intelligent system requires a "multi-memory" architecture, meaning the possession of a number of specialized yet interconnected knowledge types, including: declarative, procedural, attentional, sensory, episodic and intentional (goal-related). These knowledge types may be viewed as different sorts of patterns that a system recognizes in itself and its environment. Knowledge of these various different types must be interlinked, and in some cases may represent differing views of the same content (see Figure **??**)
3. Such a system must possess knowledge creation (i.e. pattern recognition / formation) mechanisms corresponding to each of these memory types. These mechanisms are also called "cognitive processes."
4. Each of these cognitive processes, to be effective, must have the capability to recognize when it lacks the information to perform effectively on its own; and in this case, to dynamically and interactively draw information from knowledge creation mechanisms dealing with other types of knowledge
5. This cross-mechanism interaction must have the result of enabling the knowledge creation mechanisms to perform much more effectively in combination than they would if operated non-interactively. This is "cognitive synergy."

While these points are implicit in the theory of mind given in [Goe06a], they are not articulated in this specific form there.

Interactions as mentioned in Points 4 and 5 in the above list are the real conceptual meat of the cognitive synergy idea. One way to express the key idea here is that most AI algorithms suffer from combinatorial explosions: the number of possible elements to be combined in a

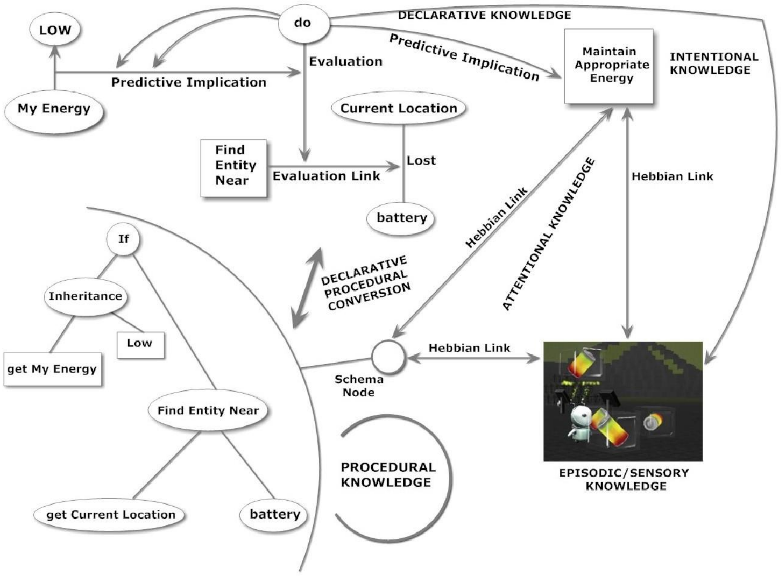

LOW
do
DECLARATIVE KNOWLEDGE
Predictive Implication
Evaluation
Predictive Implication
Maintain Appropriate Energy
INTENTIONAL KNOWLEDGE
My Energy
Current Location
Find Entity Near
Evaluation Link
Lost
battery
Hebbian Link
ATTENTIONAL KNOWLEDGE
Hebbian Link
If
DECLARATIVE PROCEDURAL CONVERSION
Inheritance
Low
get My Energy
Schema Node
Hebbian Link
Find Entity Near
PROCEDURAL KNOWLEDGE
EPISODIC/SENSORY KNOWLEDGE
get Current Location
battery

Fig. 8.1: Illustrative example of the interactions between multiple types of knowledge, in representing a simple piece of knowledge. Generally speaking, one type of knowledge can be converted to another, at the cost of some loss of information. The synergy between cognitive processes associated with corresponding pieces of knowledge, possessing different type, is a critical aspect of general intelligence.

synthesis or analysis is just too great, and the algorithms are unable to filter through all the possibilities, given the lack of intrinsic constraint that comes along with a "general intelligence" context (as opposed to a narrow-AI problem like chess-playing, where the context is constrained and hence restricts the scope of possible combinations that needs to be considered). In an AGI architecture based on cognitive synergy, the different learning mechanisms must be designed specifically to interact in such a way as to palliate each others' combinatorial explosions - so that, for instance, each learning mechanism dealing with a certain sort of knowledge, must synergize with learning mechanisms dealing with the other sorts of knowledge, in a way that decreases the severity of combinatorial explosion.

One prerequisite for cognitive synergy to work is that each learning mechanism must recognize when it is "stuck," meaning it's in a situation where it has inadequate information to make a confident judgment about what steps to take next. Then, when it does recognize that it's stuck, it may request help from other, complementary cognitive mechanisms.

A theoretical notion closely related to cognitive synergy is the *cognitive schematic*, formalized in Chapter 7 above, which states that the activity of the different cognitive processes involved in an intelligent system may be modeled in terms of the schematic implication

$$Context \wedge Procedure \rightarrow Goal$$

where the Context involves sensory, episodic and/or declarative knowledge; and attentional knowledge is used to regulate how much resource is given to each such schematic implication in memory. Synergy among the learning processes dealing with the context, the procedure and the goal is critical to the adequate execution of the cognitive schematic using feasible computational resources.

Finally, drilling a little deeper into Point 3 above, one arrives at a number of possible knowledge creation mechanisms (cognitive processes) corresponding to each of the key types of knowledge. Figure **??** below gives a high-level overview of the main types of cognitive process considered in the current version of Cognitive Synergy Theory, categorized according to the type of knowledge with which each process deals.

8.3 Cognitive Synergy in CogPrime

Different cognitive systems will use different processes to fulfill the various roles identified in Figure **??** above. Here we briefly preview the basic cognitive processes that the CogPrime AGI design uses for these roles, and the synergies that exist between these.

8.3.1 Cognitive Processes in CogPrime

: a Cognitive Synergy Based Architecture..." from ICCI 2009

Table 8.1: default

|Table will go here|

Table 8.2: The OpenCogPrime data structures used to represent the key knowledge types involved

Table 8.3: default

|Table will go here|

Table 8.4: Key cognitive processes, and the algorithms that play their roles in CogPrime

Tables 8.1 and 8.3 present the key structures and processes involved in CogPrime, identifying each one with a certain memory/process type as considered in cognitive synergy theory. That is: each of these cognitive structures or processes deals with one or more types of memory – declarative, procedural, sensory, episodic or attentional. Table 8.5 describes the key CogPrime

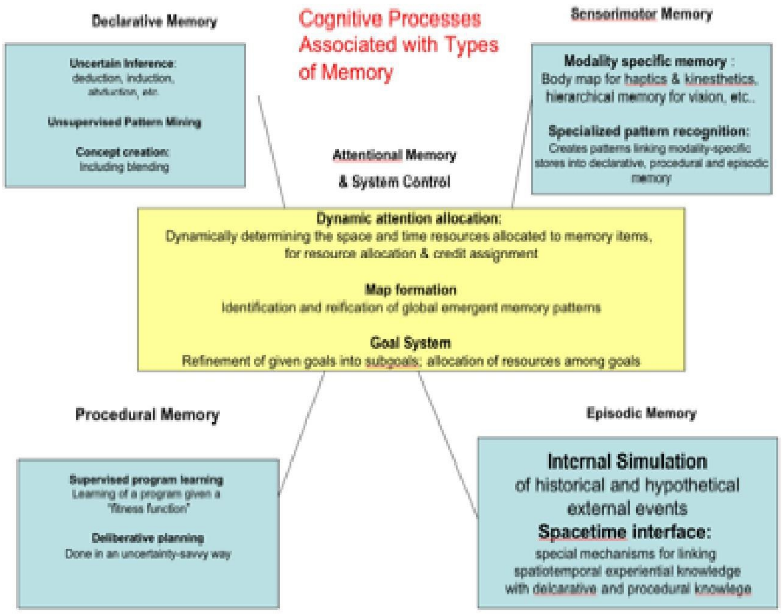

Declarative Memory
Cognitive Processes Associated with Types of Memory
Sensorimotor Memory
Uncertain Inference: deduction, induction, abduction, etc.
Unsupervised Pattern Mining
Concept creation: including blending
Modality specific memory : Body map for haptics & kinesthetics, hierarchical memory for vision, etc..
Specialized pattern recognition: Creates patterns linking modality-specific stores into declarative, procedural and episodic memory
Attentional Memory & System Control
Dynamic attention allocation: Dynamically determining the space and time resources allocated to memory items, for resource allocation & credit assignment
Map formation Identification and reification of global emergent memory patterns
Goal System Refinement of given goals into subgoals; allocation of resources among goals
Procedural Memory
Episodic Memory
Supervised program learning Learning of a program given a "fitness function"
Deliberative planning Done in an uncertainty-savvy way
Internal Simulation of historical and hypothetical external events
Spacetime interface: special mechanisms for linking spatiotemporal experiential knowledge with delcarative and procedural knowlege

Fig. 8.2: High-level overview of the key cognitive dynamics considered here in the context of cognitive synergy. The cognitive synergy principle describes the behavior of a system as it pursues a set of goals (which in most cases may be assumed to be supplied to the system “a priori”, but then refined by inference and other processes). The assumed intelligent agent model is roughly as follows: At each time the system chooses a set of procedures to execute, based on its judgments regarding which procedures will best help it achieve its goals in the current context. These procedures may involve external actions (e.g. involving conversation, or controlling an agent in a simulated world) and/or internal cognitive actions. In order to make these judgments it must effectively manage declarative, procedural, episodic, sensory and attentional memory, each of which is associated with specific algorithms and structures as depicted in the diagram. There are also global processes spanning all the forms of memory, including the allocation of attention to different memory items and cognitive processes, and the identification and reification of system-wide activity patterns (the latter referred to as “map formation”)

Table 8.5: default

|Table will go here|

Table 8.6: Key OpenCogPrime cognitive processes categorized according to knowledge type and process type

processes in terms of the "analysis vs. synthesis" distinction. Finally, Tables **??** and **??** exemplify these structures and processes in the context of embodied virtual agent control.

In the CogPrime context, a procedure in this cognitive schematic is a program tree stored in the system's procedural knowledge base; and a context is a (fuzzy, probabilistic) logical predicate stored in the AtomSpace, that holds, to a certain extent, during each interval of time. A goal is a fuzzy logical predicate that has a certain value at each interval of time, as well.

Attentional knowledge is handled in CogPrime by the ECAN artificial economics mechanism, that continually updates ShortTermImportance and LongTerm Importance values associated with each item in the CogPrime system's memory, which control the amount of attention other cognitive mechanisms pay to the item, and how much motive the system has to keep the item in memory. HebbianLinks are then created between knowledge items that often possess ShortTermImportance at the same time; this is CogPrime's version of traditional Hebbian learning.

ECAN has deep interactions with other cognitive mechanisms as well, which are essential to its efficient operation; for instance, PLN inference may be used to help ECAN extrapolate conclusions about what is worth paying attention to, and MOSES may be used to recognize subtle attentional patterns. ECAN also handles "assignment of credit", the figuring-out of the causes of an instance of successful goal-achievement, drawing on PLN and MOSES as needed when the causal inference involved here becomes difficult.

The synergies between CogPrime's cognitive processes are well summarized below, which is a 16x16 matrix summarizing a host of interprocess interactions generic to CST.

One key aspect of how CogPrime implements cognitive synergy is PLN's sophisticated management of the confidence of judgments. This ties in with the way OpenCogPrime's PLN inference framework represents truth values in terms of multiple components (as opposed to the single probability values used in many probabilistic inference systems and formalisms): each item in OpenCogPrime's declarative memory has a confidence value associated with it, which tells how much weight the system places on its knowledge about that memory item. This assists with cognitive synergy as follows: A learning mechanism may consider itself "stuck", generally speaking, when it has no high-confidence estimates about the next step it should take.

Without reasonably accurate confidence assessment to guide it, inter-component interaction could easily lead to increased rather than decreased combinatorial explosion. And of course there is an added recursion here, in that confidence assessment is carried out partly via PLN inference, which in itself relies upon these same synergies for its effective operation.

To illustrate this point further, consider one of the synergetic aspects described in **??** below: the role cognitive synergy plays in deductive inference. Deductive inference is a hard problem in general - but what is hard about it is not carrying out inference steps, but rather "inference control" (i.e., choosing which inference steps to carry out). Specifically, what must happen for deduction to succeed in CogPrime is:

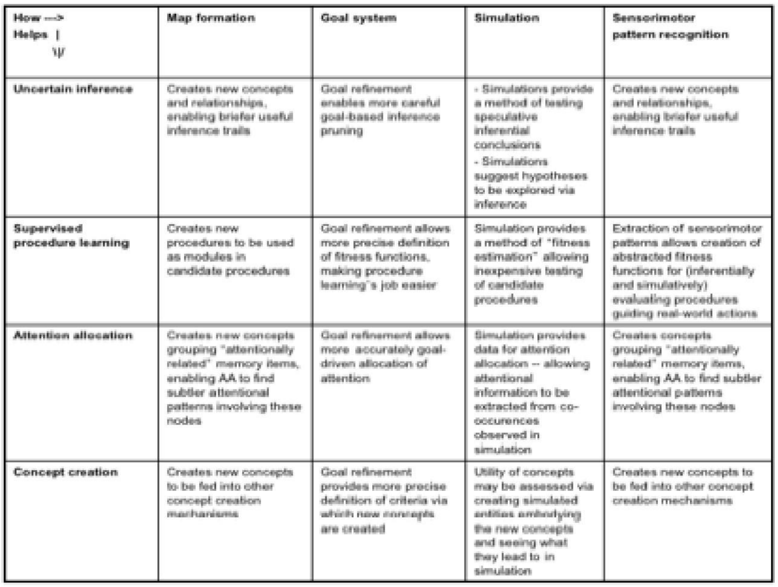
1. the system must recognize when its deductive inference process is "stuck", i.e. when the PLN inference control mechanism carrying out deduction has no clear idea regarding which inference step(s) to take next, even after considering all the domain knowledge at is disposal
2. in this case, the system must defer to another learning mechanism to gather more information about the different choices available - and the other learning mechanism chosen must, a reasonable percentage of the time, actually provide useful information that helps PLN to get "unstuck" and continue the deductive process

For instance, deduction might defer to the "attentional knowledge" subsystem, and make a judgment as to which of the many possible next deductive steps are most associated with the goal of inference and the inference steps taken so far, according to the HebbianLinks constructed by the attention allocation subsystem, based on observed associations. Or, if this fails, deduction might ask MOSES (running in supervised categorization mode) to learn predicates characterizing some of the terms involving the possible next inference steps. Once MOSES provides these new predicates, deduction can then attempt to incorporate these into its inference process, hopefully (though not necessarily) arriving at a higher-confidence next step.

8.4 Some Critical Synergies

Referring back to Figure **??**, and summarizing many of the ideas in the previous section, Table **??** enumerates a number of specific ways in which the cognitive processes mentioned in the Figure may synergize with one another, potentially achieving dramatically greater efficiency than would be possible on their own.

Of course, realizing these synergies on the practical algorithmic level requires significant inventiveness and may be approached in many different ways. The specifics of how CogPrime manifests these synergies are discussed in many following chapters.

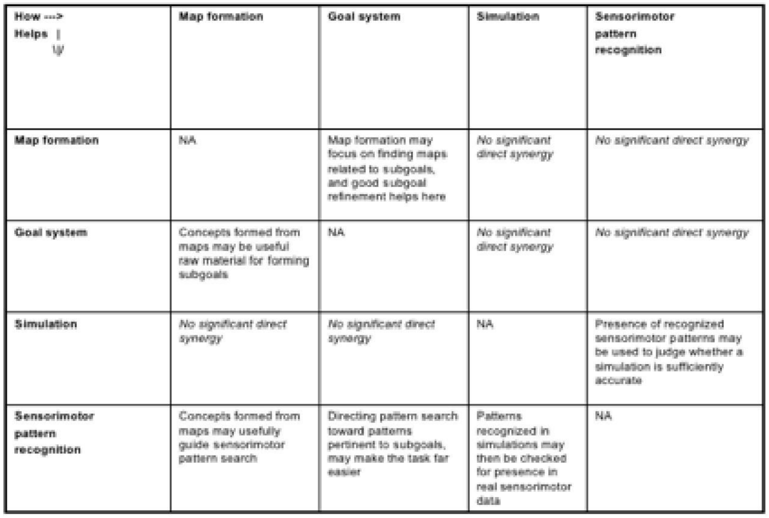


How ---> Helps | \\\|/	**Map formation**	**Goal system**	**Simulation**	**Sensorimotor pattern recognition**
Uncertain inference	Creates new concepts and relationships, enabling briefer useful inference trails	Goal refinement enables more careful goal-based inference pruning	- Simulations provide a method of testing speculative inferential conclusions - Simulations suggest hypotheses to be explored via inference	Creates new concepts and relationships, enabling briefer useful inference trails
Supervised procedure learning	Creates new procedures to be used as modules in candidate procedures	Goal refinement allows more precise definition of fitness functions, making procedure learning`s job easier	Simulation provides a method of "fitness estimation" allowing inexpensive testing of candidate procedures	Extraction of sensorimotor patterns allows creation of abstracted fitness functions for (inferentially and simulatively) evaluating procedures guiding real-world actions
Attention allocation	Creates new concepts grouping "attentionally related" memory items, enabling AA to find subtler attentional patterns involving these nodes	Goal refinement allows more accurately goal-driven allocation of attention	Simulation provides data for attention allocation -- allowing attentional information to be extracted from co-occurences observed in simulation	Creates concepts grouping "attentionally related" memory items, enabling AA to find subtler attentional patterns involving these nodes
Concept creation	Creates new concepts to be fed into other concept creation mechanisms	Goal refinement provides more precise definition of criteria via which new concepts are created	Utility of concepts may be assessed via creating simulated entities embodying the new concepts and seeing what they lead to in simulation	Creates new concepts to be fed into other concept creation mechanisms

Fig. 8.3: This table, and the following ones, show some of the synergies between the primary cognitive processes explicitly used in CogPrime.

How ---> Helps | \\|/	**Uncertain inference**	**Supervised procedure learning**	**Attention allocation**	**Concept creation**
Uncertain inference	NA	When inference gets stuck in an inference trail, it can ask procedure learning to learn new patterns regarding concepts in the inference trail (if there is adequate data regarding the concepts)	Importance levels allow pruning of inference trees	Provides new concepts, allowing briefer useful inference trails
Supervised procedure learning	Inference can be used to allow prior experience to guide each instance of procedure learning.	NA	Importance levels may be used to bias choices made in the course of procedure learning (e.g. in OCP, in the fitness evaluation and representation-building phases of MOSES)	Provides new concepts, allowing compacter programs using new concepts in various roles
Attention allocation	Enables inference of new HebbianLinks and HebbianPredicates from existing ones	Procedure learning can recognize patterns in historical system activity, which are then used to build concepts and relationships guiding attention allocation	NA	Combination of concepts formed via map formation, may lead to new concepts that even better direct attention
Concept creation	Allows inferential assessment of the value of new concepts	Procedure learning can be used to search for high-quality blends of existing concepts (using e.g. inferential and attentional knowledge in the fitness functions)	Allows assessment of the value of new concepts based on historical attentional knowledge	NA

How ---> Helps | \\|/	**Uncertain inference**	**Supervised procedure learning**	**Attention allocation**	**Concept creation**
Map formation	Speculative inference can help map formation guess which maps to hunt for	Procedure learning can be used to search for maps that are more complex than mere "co-occurrence"	Attention allocation provides the raw data for map formation	*No significant direct synergy*
Goal system	Inference can carry out goal refinement	*No significant direct synergy*	Flow of importance among subgoals determines which subgoals get used, versus being forgotten	Concept creation can be used to provide raw data for goal refinement (e.g. a new subgoal that blends two others)
Simulation	In order to provide data for setting up simulations, inference will often be needed	*No significant direct synergy*	Attention allocation tells which portions of a simulation need to be run in more detail	*No significant direct synergy*
Sensorimotor pattern recognition	Speculative inference helps fill in gaps in sensory data	Procedure learning can be used to find subtle patterns in sensorimotor data	Attention allocation guides pattern recognition via indicating which sensorimotor stimuli and patterns tend to be associatively linked	New concepts may be created that then are found to serve as significant patterns in sensorimotor data



How ---> **Helps |** \\/	**Map formation**	**Goal system**	**Simulation**	**Sensorimotor pattern recognition**
Map formation	NA	Map formation may focus on finding maps related to subgoals, and good subgoal refinement helps here	*No significant direct synergy*	*No significant direct synergy*
Goal system	Concepts formed from maps may be useful raw material for forming subgoals	NA	*No significant direct synergy*	*No significant direct synergy*
Simulation	*No significant direct synergy*	*No significant direct synergy*	NA	Presence of recognized sensorimotor patterns may be used to judge whether a simulation is sufficiently accurate
Sensorimotor pattern recognition	Concepts formed from maps may usefully guide sensorimotor pattern search	Directing pattern search toward patterns pertinent to subgoals, may make the task far easier	Patterns recognized in simulations may then be checked for presence in real sensorimotor data	NA

8.5 The Cognitive Schematic

Now we return to the "cognitive schematic" notion, according to which various cognitive processes involved in intelligence may be understood to work together via the implication

$$Context \wedge Procedure \rightarrow Goal < p >$$

(summarized $C \wedge P \rightarrow G$). Semi-formally, this implication may be interpreted to mean: "If the context C appears to hold currently, then if I enact the procedure P, I can expect to achieve the goal G with certainty p."

The cognitive schematic leads to a conceptualization of the internal action of an intelligent system as involving two key categories of learning:

- **Analysis**: Estimating the probability p of a posited $C \wedge P \rightarrow G$ relationship
- **Synthesis**: Filling in one or two of the variables in the cognitive schematic, given assumptions regarding the remaining variables, and directed by the goal of maximizing the probability of the cognitive schematic

More specifically, where synthesis is concerned, some key examples are:

- The MOSES probabilistic evolutionary program learning algorithm is applied to find P, given fixed C and G. Internal simulation is also used, for the purpose of creating a simulation embodying C and seeing which P lead to the simulated achievement of G.
  - *Example: A virtual dog learns a procedure P to please its owner (the goal G) in the context C where there is a ball or stick present and the owner is saying "fetch".*
- PLN inference, acting on declarative knowledge, is used for choosing C, given fixed P and G (also incorporating sensory and episodic knowledge as appropriate). Simulation may also be used for this purpose.

  - *Example: A virtual dog wants to achieve the goal G of getting food, and it knows that the procedure P of begging has been successful at this before, so it seeks a context C where begging can be expected to get it food. Probably this will be a context involving a friendly person.*
- PLN-based goal refinement is used to create new subgoals G to sit on the right hand side of instances of the cognitive schematic.
  - *Example: Given that a virtual dog has a goal of finding food, it may learn a subgoal of following other dogs, due to observing that other dogs are often heading toward their food.*
- Concept formation heuristics are used for choosing G and for fueling goal refinement, but especially for choosing C (via providing new candidates for C). They are also used for choosing P, via a process called "predicate schematization" that turns logical predicates (declarative knowledge) into procedures.
  - *Example: At first a virtual dog may have a hard time predicting which other dogs are going to be mean to it. But it may eventually observe common features among a number of mean dogs, and thus form its own concept of "pit bull," without anyone ever teaching it this concept explicitly.*

Where analysis is concerned:

- PLN inference, acting on declarative knowledge, is used for estimating the probability of the implication in the cognitive schematic, given fixed C, P and G. Episodic knowledge is also used this regard, via enabling estimation of the probability via simple similarity matching against past experience. Simulation is also used: multiple simulations may be run, and statistics may be captured therefrom.
  - *Example: To estimate the degree to which asking Bob for food (the procedure P is "asking for food", the context C is "being with Bob") will achieve the goal G of getting food, the virtual dog may study its memory to see what happened on previous occasions where it or other dogs asked Bob for food or other things, and then integrate the evidence from these occasions.*
- Procedural knowledge, mapped into declarative knowledge and then acted on by PLN inference, can be useful for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C \wedge P_1 \rightarrow G$ is known for some P_1 related to P.
  - *Example: knowledge of the internal similarity between the procedure of asking for food and the procedure of asking for toys, allows the virtual dog to reason that if asking Bob for toys has been successful, maybe asking Bob for food will be successful too.*
- Inference, acting on declarative or sensory knowledge, can be useful for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C_1 \wedge P \rightarrow G$ is known for some C_1 related to C.
  - *Example: if Bob and Jim have a lot of features in common, and Bob often responds positively when asked for food, then maybe Jim will too.*
- Inference can be used similarly for estimating the probability of the implication $C \wedge P \rightarrow G$, in cases where the probability of $C \wedge P \rightarrow G_1$ is known for some G_1 related to G. Concept

creation can be useful indirectly in calculating these probability estimates, via providing new concepts that can be used to make useful inference trails more compact and hence easier to construct.

  - *Example: The dog may reason that because Jack likes to play, and Jack and Jill are both children, maybe Jill likes to play too. It can carry out this reasoning only if its concept creation process has invented the concept of "child" via analysis of observed data.*

In these examples we have focused on cases where two terms in the cognitive schematic are fixed and the third must be filled in; but just as often, the situation is that only one of the terms is fixed. For instance, if we fix G, sometimes the best approach will be to collectively learn C and P. This requires either a procedure learning method that works interactively with a declarative-knowledge-focused concept learning or reasoning method; or a declarative learning method that works interactively with a procedure learning method. That is, it requires the sort of cognitive synergy built into the CogPrime design.

8.6 Cognitive Synergy for Procedural and Declarative Learning

We now present a little more algorithmic detail regarding the operation and synergetic interaction of CogPrime's two most sophisticated components: the MOSES procedure learning algorithm (see Chapter 33), and the PLN uncertain inference framework (see Chapter 34). The treatment is necessarily quite compact, since we have not yet reviewed the details of either MOSES or PLN; but as well as illustrating the notion of cognitive synergy more concretely, perhaps the high-level discussion here will make clearer how MOSES and PLN fit into the big picture of CogPrime.

8.6.1 Cognitive Synergy in MOSES

MOSES, CogPrime's primary algorithm for learning procedural knowledge, has been tested on a variety of application problems including standard GP test problems, virtual agent control, biological data analysis and text classification [Loo06]. It represents procedures internally as program trees. Each node in a MOSES program tree is supplied with a "knob," comprising a set of values that may potentially be chosen to replace the data item or operator at that node. So for instance a node containing the number 7 may be supplied with a knob that can take on any integer value. A node containing a while loop may be supplied with a knob that can take on various possible control flow operators including conditionals or the identity. A node containing a procedure representing a particular robot movement, may be supplied with a knob that can take on values corresponding to multiple possible movements. Following a metaphor suggested by Douglas Hofstadter [Hof96], MOSES learning covers both "knob twiddling" (setting the values of knobs) and "knob creation."

MOSES is invoked within CogPrime in a number of ways, but most commonly for finding a procedure P satisfying a probabilistic implication $C \& P \rightarrow G$ as described above, where C is an observed context and G is a system goal. In this case the probability value of the implication provides the "scoring function" that MOSES uses to assess the quality of candidate procedures.

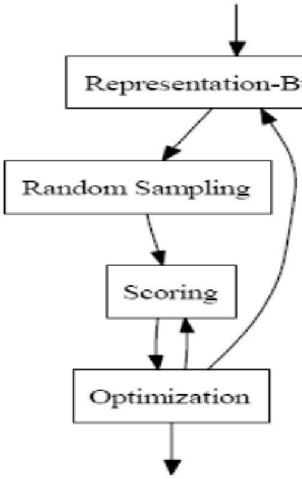


Fig. 8.4: High-Level Control Flow of MOSES Algorithm

For example, suppose an CogPrime -controlled robot is trying to learn to play the game of “tag." (I.e. a multi-agent game in which one agent is specially labeled "it", and runs after the other player agents, trying to touch them. Once another agent is touched, it becomes the new "it" and the previous "it" becomes just another player agent.) Then its context C is that others are trying to play a game they call “tag” with it; and we may assume its goals are to please them and itself, and that it has figured out that in order to achieve this goal it should learn some procedure to follow when interacting with others who have said they are playing “tag.” In this case a potential tag-playing procedure might contain nodes for physical actions like $step\_forward(speed\ s)$, as well as control flow nodes containing operators like $ifelse$ (for instance, there would probably be a conditional telling the robot to do something different depending on whether someone seems to be chasing it). Each of these program tree nodes would have an appropriate knob assigned to it. And the scoring function would evaluate a procedure P in terms of how successfully the robot played tag when controlling its behaviors according to P (noting that it may also be using other control procedures concurrently with P). It’s worth noting here that evaluating the scoring function in this case involves some inference already, because in order to tell if it is playing tag successfully, in a real-world context, it must watch and understand the behavior of the other players.

MOSES follows the high-level control flow depicted in Figure 8.4, which corresponds to the following process for evolving a metapopulation of “demes“ of programs (each deme being a set of relatively similar programs, forming a sort of island in program space):

1. Construct an initial set of knobs based on some prior (e.g., based on an empty program; or more interestingly, using prior knowledge **supplied by PLN inference** based on the system’s memory) and use it to generate an initial random sampling of programs. Add this deme to the metapopulation.
2. Select a deme from the metapopulation and update its sample, as follows:

   a. Select some promising programs from the deme's existing sample to use for modeling, according to the scoring function.
   b. Considering the promising programs as collections of knob settings, generate new collections of knob settings by applying some (competent) optimization algorithm. For best performance on difficult problems, it is important to use an optimization algorithm that makes use of the system's memory in its choices, **consulting PLN inference** to help estimate which collections of knob settings will work best.
   c. Convert the new collections of knob settings into their corresponding programs, reduce the programs to normal form, evaluate their scores, and integrate them into the deme's sample, replacing less promising programs. In the case that scoring is expensive, score evaluation may be preceded by score estimation, which may use **PLN inference**, enaction of procedures in an **internal simulation environment**, and/or similarity matching against **episodic memory**.

3. For each new program that meet the criterion for creating a new deme, if any:

   a. Construct a new set of knobs (a process called "representation-building") to define a region centered around the program (the deme's *exemplar)*, and use it to generate a *new* random sampling of programs, producing a new deme.
   b. Integrate the new deme into the metapopulation, possibly displacing less promising demes.

4. Repeat from step 2.

MOSES is a complex algorithm and each part plays its role; if any one part is removed the performance suffers significantly [Loo06]. However, the main point we want to highlight here is the role played by synergetic interactions between MOSES and other cognitive components such as PLN, simulation and episodic memory, as indicated in **boldface** in the above pseudocode. MOSES is a powerful procedure learning algorithm, but used on its own it runs into scalability problems like any other such algorithm; the reason we feel it has potential to play a major role in a human-level AI system is its capacity for productive interoperation with other cognitive components.

Continuing the "tag" example, the power of MOSES's integration with other cognitive processes would come into play if, before learning to play tag, the robot has already played simpler games involving chasing. If the robot already has experience chasing and being chased by other agents, then its episodic and declarative memory will contain knowledge about how to pursue and avoid other agents in the context of running around an environment full of objects, and this knowledge will be deployable within the appropriate parts of MOSES's Steps 1 and 2. Cross-process and cross-memory-type integration make it tractable for MOSES to act as a "transfer learning" algorithm, not just a task-specific machine-learning algorithm.

8.6.2 Cognitive Synergy in PLN

While MOSES handles much of CogPrime's procedural learning, and OpenCogPrimes internal simulation engine handles most episodic knowledge, CogPrime's primary tool for handling declarative knowledge is an uncertain inference framework called Probabilistic Logic Networks (PLN). The complexities of PLN are the topic of a lengthy technical monograph [GMIH08], and here we will eschew most details and focus mainly on pointing out how PLN seeks to achieve efficient inference control via integration with other cognitive processes.

As a logic, PLN is broadly integrative: it combines certain term logic rules with more standard predicate logic rules, and utilizes both fuzzy truth values and a variant of imprecise probabilities called *indefinite probabilities*. PLN mathematics tells how these uncertain truth values propagate through its logic rules, so that uncertain premises give rise to conclusions with reasonably accurately estimated uncertainty values. This careful management of uncertainty is critical for the application of logical inference in the robotics context, where most knowledge is abstracted from experience and is hence highly uncertain.

PLN can be used in either forward or backward chaining mode; and in the language introduced above, it can be used for either analysis or synthesis. As an example, we will consider backward chaining analysis, exemplified by the problem of a robot preschool-student trying to determine whether a new playmate "Bob" is likely to be a regular visitor to its preschool or not (evaluating the truth value of the implication $Bob \rightarrow regular\_visitor$). The basic backward chaining process for PLN analysis looks like:

1. Given an implication $L \equiv A \rightarrow B$ whose truth value must be estimated (for instance $L \equiv C\&P \rightarrow G$ as discussed above), create a list $(A_1, ..., A_n)$ of *(inference rule, stored knowledge)* pairs that might be used to produce L
2. Using analogical reasoning to prior inferences, assign each A_i a probability of success
    - If some of the A_i are estimated to have reasonable probability of success at generating reasonably confident estimates of L's truth value, then invoke Step 1 with A_i in place of L (at this point the inference process becomes recursive)
    - If none of the A_i looks sufficiently likely to succeed, then inference has "gotten stuck" and another cognitive process should be invoked, e.g.
        - **Concept creation** may be used to infer new concepts related to A and B, and then Step 1 may be revisited, in the hope of finding a new, more promising A_i involving one of the new concepts
        - **MOSES** may be invoked with one of several special goals, e.g. the goal of finding a procedure P so that $P(X)$ predicts whether $X \rightarrow B$. If MOSES finds such a procedure P then this can be converted to declarative knowledge understandable by PLN and Step 1 may be revisited....
        - **Simulations** may be run in CogPrime's internal simulation engine, so as to observe the truth value of $A \rightarrow B$ in the simulations; and then Step 1 may be revisited....

The combinatorial explosion of inference control is combatted by the capability to defer to other cognitive processes when the inference control procedure is unable to make a sufficiently confident choice of which inference steps to take next. Note that just as MOSES may rely on PLN to model its evolving populations of procedures, PLN may rely on MOSES to create complex knowledge about the terms in its logical implications. This is just one example of the multiple ways in which the different cognitive processes in CogPrime interact synergetically; a more thorough treatment of these interactions is given in Chapter 49.

In the "new playmate" example, the interesting case is where the robot initially seems not to know enough about Bob to make a solid inferential judgment (so that none of the A_i seem particularly promising). For instance, it might carry out a number of possible inferences and not come to any reasonably confident conclusion, so that the reason none of the A_i seem promising is that all the decent-looking ones have been tried already. So it might then recourse to MOSES, simulation or concept creation.

For instance, the PLN controller could make a list of everyone who has been a regular visitor, and everyone who has not been, and pose MOSES the task of figuring out a procedure for distinguishing these two categories. This procedure could then used directly to make the needed assessment, or else be translated into logical rules to be used within PLN inference. For example, perhaps MOSES would discover that older males wearing ties tend not to become regular visitors. If the new playmate is an older male wearing a tie, this is directly applicable. But if the current playmate is wearing a tuxedo, then PLN may be helpful via reasoning that even though a tuxedo is not a tie, it's a similar form of fancy dress – so PLN may extend the MOSES-learned rule to the present case and infer that the new playmate is not likely to be a regular visitor.

8.7 Is Cognitive Synergy Tricky?

[1]

In this section we use the notion of cognitive synergy to explore a question that arises frequently in the AGI community: the well-known difficulty of measuring intermediate progress toward human-level AGI. We explore some potential reasons underlying this, via extending the notion of cognitive synergy to a more refined notion of "tricky cognitive synergy." These ideas are particularly relevant to the problem of creating a roadmap toward AGI, as we'll explore in Chapter 17 below.

8.7.1 The Puzzle: Why Is It So Hard to Measure Partial Progress Toward Human-Level AGI?

It's not entirely straightforward to create tests to measure the *final achievement* of human-level AGI, but there are some fairly obvious candidates here. There's the Turing Test (fooling judges into believing you're human, in a text chat), the video Turing Test, the Robot College Student test (passing university, via being judged exactly the same way a human student would), etc. There's certainly no agreement on which is the most meaningful such goal to strive for, but there's broad agreement that a number of goals of this nature basically make sense.

On the other hand, how does one measure whether one is, say, 50 percent of the way to human-level AGI? Or, say, 75 or 25 percent?

It's possible to pose many "practical tests" of incremental progress toward human-level AGI, with the property that if a proto-AGI system passes the test using a certain sort of architecture and/or dynamics, then this implies a certain amount of progress toward human-level AGI *based on particular theoretical assumptions about AGI.* However, in each case of such a practical test, it seems intuitively likely *to a significant percentage of AGI researchers* that there is some way to "game" the test via designing a system specifically oriented toward passing that test, and which doesn't constitute dramatic progress toward AGI.

Some examples of practical tests of this nature would be

---

[1] This section co-authored with Jared Wigmore

- The Wozniak "coffee test": go into an average American house and figure out how to make coffee, including identifying the coffee machine, figuring out what the buttons do, finding the coffee in the cabinet, etc.
- Story understanding – reading a story, or watching it on video, and then answering questions about what happened (including questions at various levels of abstraction)
- Graduating (virtual-world or robotic) preschool
- Passing the elementary school reading curriculum (which involves reading and answering questions about some picture books as well as purely textual ones)
- Learning to play an arbitrary video game based on experience only, or based on experience plus reading instructions

One interesting point about tests like this is that each of them seems to *some* AGI researchers to encapsulate the crux of the AGI problem, and be unsolvable by any system not far along the path to human-level AGI – yet seems to other AGI researchers, with different conceptual perspectives, to be something probably game-able by narrow-AI methods. And of course, given the current state of science, there's no way to tell which of these practical tests really can be solved via a narrow-AI approach, except by having a lot of people try really hard over a long period of time.

A question raised by these observations is whether there is some *fundamental reason* why it's hard to make an objective, theory-independent measure of intermediate progress toward advanced AGI. Is it just that we haven't been smart enough to figure out the right test – or is there some conceptual reason why the very notion of such a test is problematic?

We don't claim to know for sure – but in the rest of this section we'll outline one possible reason why the latter might be the case.

8.7.2 A Possible Answer: Cognitive Synergy is Tricky!

Why might a solid, objective empirical test for intermediate progress toward AGI be an infeasible notion? One possible reason, we suggest, is precisely *cognitive synergy*, as discussed above.

The cognitive synergy hypothesis, in its simplest form, states that human-level AGI intrinsically depends on the synergetic interaction of multiple components (for instance, as in CogPrime, multiple memory systems each supplied with its own learning process). In this hypothesis, for instance, it might be that there are 10 critical components required for a human-level AGI system. Having all 10 of them in place results in human-level AGI, but having only 8 of them in place results in having a dramatically impaired system – and maybe having only 6 or 7 of them in place results in a system that can hardly do anything at all.

Of course, the reality is almost surely not as strict as the simplified example in the above paragraph suggests. No AGI theorist has really posited a list of 10 crisply-defined subsystems and claimed them necessary and sufficient for AGI. We suspect there are many different routes to AGI, involving integration of different sorts of subsystems. However, if the cognitive synergy hypothesis is correct, then human-level AGI behaves *roughly* like the simplistic example in the prior paragraph suggests. Perhaps instead of using the 10 components, you could achieve human-level AGI with 7 components, but having only 5 of these 7 would yield drastically impaired functionality – etc. Or the point could be made without any decomposition into a finite set of components, using continuous probability distributions. To mathematically formalize the cognitive synergy hypothesis becomes complex, but here we're only aiming for a qualitative argument. So for illustrative purposes, we'll stick with the "10 components" example, just for communicative simplicity.

Next, let's suppose that for any given task, there are ways to achieve this task using a system that is much simpler than any subset of size 6 drawn from the set of 10 components needed for human-level AGI, but works much better for the task than this subset of 6 components (assuming the latter are used as a set of only 6 components, without the other 4 components).

Note that this supposition is a good bit stronger than mere cognitive synergy. For lack of a better name, we'll call it *tricky cognitive synergy*. The tricky cognitive synergy hypothesis would be true if, for example, the following possibilities were true:

- creating components to serve as parts of a synergetic AGI is *harder* than creating components intended to serve as parts of simpler AI systems without synergetic dynamics
- components capable of serving as parts of a synergetic AGI are necessarily *more complicated* than components intended to serve as parts of simpler AGI systems.

These certainly seem reasonable possibilities, since to serve as a component of a synergetic AGI system, a component must have the internal flexibility to usefully handle interactions with a lot of other components as well as to solve the problems that come its way. In a CogPrime context, these possibilities ring true, in the sense that tailoring an AI process for tight integration with other AI processes within CogPrime, tends to require more work than preparing a conceptually similar AI process for use on its own or in a more task-specific narrow AI system.

It seems fairly obvious that, if tricky cognitive synergy really holds up as a property of human-level general intelligence, the difficulty of formulating tests for intermediate progress toward human-level AGI follows as a consequence. Because, according to the tricky cognitive synergy hypothesis, any test is going to be more easily solved by some simpler narrow AI process than by a *partially complete* human-level AGI system.

8.7.3 Conclusion

We haven't proved anything here, only made some qualitative arguments. However, these arguments do seem to give a plausible explanation for the empirical observation that positing tests for intermediate progress toward human-level AGI is a very difficult prospect. If the theoretical notions sketched here are correct, then this difficulty is not due to incompetence or lack of imagination on the part of the AGI community, nor due to the primitive state of the AGI field, but is rather intrinsic to the subject matter. And if these notions are correct, then quite likely the future rigorous science of AGI will contain formal theorems echoing and improving the qualitative observations and conjectures we've made here.

If the ideas sketched here are true, then the practical consequence for AGI development is, very simply, that one shouldn't worry a lot about producing intermediary results that are compelling to skeptical observers. Just at 2/3 of a human brain may not be of much use, similarly, 2/3 of an AGI system may not be much use. Lack of impressive intermediary results may not imply one is on a wrong development path; and comparison with narrow AI systems on specific tasks may be badly misleading as a gauge of incremental progress toward human-level AGI.

Hopefully it's clear that the motivation behind the line of thinking presented here is a desire to understand the nature of general intelligence and its pursuit – not a desire to avoid testing our AGI software! Really, as AGI engineers, we would love to have a sensible rigorous way to test our intermediary progress toward AGI, so as to be able to pose convincing arguments to skeptics, funding sources, potential collaborators and so forth. Our motivation here is not a desire to avoid having the intermediate progress of our efforts measured, but rather a desire to explain the frustrating (but by now rather well-established) difficulty of creating such intermediate goals for human-level AGI in a meaningful way.

If we or someone else figures out a compelling way to measure partial progress toward AGI, we will celebrate the occasion. But it seems worth seriously considering the possibility that the difficulty in finding such a measure reflects fundamental properties of general intelligence.

From a practical CogPrime perspective, we are interested in a variety of evaluation and testing methods, including the "virtual preschool" approach mentioned briefly above and more extensively in later chapters. However, our focus will be on evaluation methods that give us meaningful information about CogPrime's progress, given our knowledge of how CogPrime works and our understanding of the underlying theory. We are unlikely to focus on the achievement of intermediate test results capable of convincing skeptics of the reality of our partial progress, because we have not yet seen any credible tests of this nature, and because we suspect the reasons for this lack may be rooted in deep properties of feasible general intelligence, such as tricky cognitive synergy.

Chapter 9
General Intelligence in the Everyday Human World

9.1 Introduction

Intelligence is not just about what happens inside a system, but also about what happens outside that system, and how the system interacts with its environment. Real-world general intelligence is about intelligence *relative to some particular class of environments*, and human-like general intelligence is about intelligence relative to the particular class of environments that humans evolved in (which in recent millennia has included environments humans have created using their intelligence). In Chapter 2, we reviewed some specific capabilities characterizing human-like general intelligence; to connect these with the general theory of general intelligence from the last few chapters, we need to explain what aspects of human-relevant *environments* correspond to these human-like intelligent *capabilities*. We begin with aspects of the environment related to communication, which turn out to tie in closely with cognitive synergy. Then we turn to physical aspects of the environment, which we suspect also connect closely with various human cognitive capabilities. Finally we turn to physical aspects of the human body and their relevance to the human mind. In the following chapter we present a deeper, more abstract theoretical framework encompassing these ideas.

These ideas are of theoretical importance, and they're also of practical importance when one turns to the critical area of *AGI environment design*. If one is going to do anything besides release one's young AGI into the "wilds" of everyday human life, then one has to put some thought into what kind of environment it will be raised in. This may be a virtual world or it may be a robot preschool or some other kind of physical environment, but in any case some specific choices must be made about what to include. Specific choices must also be made about what kind of body to give one's AGI system – what sensors and actuators, and so forth. In Chapter 16 we will present some specific suggestions regarding choices of embodiment and environment that we find to be ideal for AGI development – virtual and robot preschools – but the material in this chapter is of more general import, beyond any such particularities. If one has an intuitive idea of what properties of body and world human intelligence is biased for, then one can make practical choices about embodiment and environment in a principled rather than purely ad hoc or opportunistic way.

9.2 Some Broad Properties of the Everyday World That Help Structure Intelligence

The properties of the everyday world that help structure intelligence are diverse and span multiple levels of abstraction. Most of this chapter will focus on fairly concrete patterns of this nature, such as are involved in inter-agent communication and naive physics; however, it's also worth noting the potential importance of more abstract patterns distinguishing the everyday world from arbitrary mathematical environments.

The propensity to search for hierarchical patterns is one huge potential example of an abstract everyday-world property. We strongly suspect the reason that searching for hierarchical patterns works so well, in so many everyday-world contexts, lies in the particular structure of the everyday world – it's not something that would be true across all possible environments (even if one weights the space of possible environments in some clever way, say using program-length according to some standard computational model). However, this sort of assertion is of course highly "philosophical," and becomes complex to formulate and defend convincingly given the current state of science and mathematics.

Going one step further, we recall from Chapter 3 a structure called the "dual network", which consists of superposed hierarchical and heterarchical networks: basically a hierarchy in which the distance between two nodes in the hierarchy is correlated with the distance between the nodes in some metric space. Another high level property of the everyday world may be that dual network structures are prevalent. This would imply that minds biased to represent the world in terms of dual network structure are likely to be intelligent with respect to the everyday world.

In a different direction, the extreme commonality of symmetry groups in the (everyday and otherwise) physical world is another example: they occur so often that minds oriented toward recognizing patterns involving symmetry groups are likely to be intelligent with respect to the real world.

We suspect that the number of cognitively-relevant properties of the everyday world is huge ... and that the essence of everyday-world intelligence lies in the list of varyingly abstract and concrete properties, which must be embedded implicitly or explicitly in the structure of a natural or artificial intelligence for that system to have everyday-world intelligence.

Apart from these particular yet abstract properties of the everyday world, intelligence is just about "finding patterns in which actions tend to achieve which goals in which situations" ... but, the simple meta-algorithm needed to accomplish this universally is, we suggest, only a small percentage what it takes to make a mind.

You might say that a sufficiently generally intelligent system should be able to infer the various cognitively-relevant properties of the environment from looking at data about the everyday world. We agree *in principle*, and in fact Ben Kuipers and his colleagues have done some interesting work in this direction, showing that learning algorithms can infer some basics about the structure of space and time from experience [MK07]. But we suggest that doing this really thoroughly would require a massively greater amount of processing power than an AGI that embodies and hence automatically utilizes these principles. It may be that the problem of inferring these properties is so hard as to require a wildly infeasible $AIXI^{tl}$ / Godel Machine type system.

9.3 Embodied Communication

Next we turn to the potential cognitive implications of seeking to achieve goals in an environment in which multimodal communication with other agents plays a prominent role.

Consider a community of embodied agents living in a shared world, and suppose that the agents can communicate with each other via a set of mechanisms including:

- **Linguistic communication**, in a language whose semantics is largely (not necessarily wholly) interpretable based on the mutually experienced world
- **Indicative communication**, in which e.g. one agent points to some part of the world or delimits some interval of time, and another agent is able to interpret the meaning
- **Demonstrative communication**, in which an agent carries out a set of actions in the world, and the other agent is able to imitate these actions, or instruct another agent as to how to imitate these actions
- **Depictive communication**, in which an agent creates some sort of (visual, auditory, etc.) construction to show another agent, with a goal of causing the other agent to experience phenomena similar to what they would experience upon experiencing some particular entity in the shared environment
- **Intentional communication**, in which an agent explicitly communicates to another agent what its goal is in a certain situation [1]

It is clear that ordinary everyday communication between humans possesses all these aspects.

We define the **Embodied Communication Prior** (ECP) as the probability distribution in which the probability of an entity (e.g. a goal or environment) is proportional to the difficulty of describing that entity, for a typical member of the community in question, using a particular set of communication mechanisms including the above five modes. We will sometimes refer to the prior probability of an entity under this distribution, as its "simplicity" under the distribution.

Next, to further specialize the Embodied Communication Prior, we will assume that for each of these modes of communication, there are some aspects of the world that are much more easily communicable using that mode than the other modes. For instance, in the human everyday world:

- Abstract (declarative) statements spanning large classes of situations are generally much easier to communicate linguistically
- Complex, multi-part procedures are much easier to communicate either demonstratively, or using a combination of demonstration with other modes
- Sensory or episodic data is often much easier to communicate demonstratively
- The current value of attending to some portion of the shared environment is often much easier to communicate indicatively
- Information about what goals to follow in a certain situation is often much easier to communicate intentionally, i.e. via explicitly indicating what one's own goal is

These simple observations have significant implications for the nature of the Embodied Communication Prior. For one thing they let us define multiple forms of knowledge:

- **Isolatedly declarative knowledge** is that which is much more easily communicable linguistically

---

[1] in Appendix **??** we recount some interesting recent results showing that mirror neurons fire in response to some cases of intentional communication as thus defined

- **Isolatedly procedural knowledge** is that which is much more easily communicable demonstratively
- **Isolatedly sensory knowledge** is that which is much more easily communicable depictively
- **Isolatedly attentive knowledge** is that which is much more easily communicable indicatively
- **Isolatedly intentional knowledge** is that which is much more easily communicable intentionally

This categorization of knowledge types resembles many ideas from the cognitive theory of memory [TC05], although the distinctions drawn here are a little crisper than any classification currently derivable from available neurological or psychological data.

Of course there may be much knowledge, of relevance to systems seeking intelligence according to the ECP, that does not fall into any of these categories and constitutes "mixed knowledge." There are some very important specific subclasses of mixed knowledge. For instance, episodic knowledge (knowledge about specific real or hypothetical sets of events) will most easily be communicated via a combination of declarative, sensory and (in some cases) procedural communication. Scientific and mathematical knowledge are generally mixed knowledge, as is most everyday commonsense knowledge.

Some cases of mixed knowledge are reasonably well decomposable, in the sense that they decompose into knowledge items that individually fall into some specific knowledge type. For instance, an experimental chemistry procedure may be much more easily communicable procedurally, whereas an allied piece of knowledge from theoretical chemistry may be much more easily communicable declaratively; but in order to fully communicate either the experimental procedure or the abstract piece of knowledge, one may ultimately need to communicate both aspects.

Also, even when the best way to communicate something is mixed-mode, it may be possible to identify one mode that poses the most important part of the communication. An example would be a chemistry experiment that is best communicated via a practical demonstration together with a running narrative. It may be that the demonstration without the narrative would be vastly more valuable than the narrative without the demonstration. To cover such cases we may make less restrictive definitions such as

- **Interactively declarative knowledge** is that which is much more easily communicable in a manner dominated by linguistic communication

and so forth. We call these "interactive knowledge categories," by contrast to the "isolated knowledge categories" introduced earlier.

9.3.0.1 Naturalness of Knowledge Categories

Next we introduce an assumption we call NKC, for Naturalness of Knowledge Categories. The NKC assumption states that the knowledge in each of the above isolated and interactive communication-modality-focused categories forms a "natural category," in the sense that for each of these categories, there are many different properties shared by a large percentage of the knowledge in the category, but not by a large percentage of the knowledge in the other categories. This means that, for instance, procedural knowledge systematically (and statistically) has different characteristics than the other kinds of knowledge.

The NKC assumption seems commonsensically to hold true for human everyday knowledge, and it has fairly dramatic implications for general intelligence. Suppose we conceive general intelligence as the ability to achieve goals in the environment shared by the communicating agents underlying the Embodied Communication Prior. Then, NKC suggests that the best way to achieve general intelligence according to the Embodied Communication Prior is going to involve

- specialized methods for handling declarative, procedural, sensory and attentional knowledge (due to the naturalness of the isolated knowledge categories)
- specialized methods for handling interactions between different types of knowledge, including methods focused on the case where one type of knowledge is primary and the others are supporting (the latter due to the naturalness of the interactive knowledge categories)

9.3.0.2 Cognitive Completeness

Suppose we conceive an AI system as consisting of a set of learning capabilities, each one characterized by three features:

- One or more **knowledge types** that it is competent to deal with, in the sense of the two key learning problems mentioned above
- At least one **learning type**: either analysis, or synthesis, or both
- At least one **interaction type**, for each (knowledge type, learning type) pair it handles: "isolated" (meaning it deals mainly with that knowledge type in isolation), or "interactive" (meaning it focuses on that knowledge type but in a way that explicitly incorporates other knowledge types into its process), or "fully mixed" (meaning that when it deals with the knowledge type in question, no particular knowledge type tends to dominate the learning process).

Then, intuitively, it seems to follow from the ECP with NKC that systems with high efficient general intelligence should have the following properties, which collectively we'll call **cognitive completeness**:

- For each (knowledge type, learning type, interaction type) triple, there should be a learning capability corresponding to that triple.
- Furthermore the capabilities corresponding to different (knowledge type, interaction type) pairs should have distinct characteristics (since according to the NKC the isolated knowledge corresponding to a knowledge type is a natural category, as is the dominant knowledge corresponding to a knowledge type)
- For each (knowledge type, learning type) pair (K,L), and each other knowledge type K1 distinct from K, there should be a distinctive capability with interaction type "interactive" and dealing with knowledge that is interactively K but also includes aspects of K1

Furthermore, it seems intuitively sensible that according to the ECP with NKC, if the capabilities mentioned in the above points are reasonably able, then the system possessing the capabilities will display general intelligence relative to the ECP. Thus we arrive at the hypothesis that

Under the assumption of the Embodied Communication Prior (with the Natural Knowledge Categories assumption), the property above called "cognitive completeness" is necessary and sufficient for efficient general intelligence at the level of an inteligent adult human (e.g. at the Piagetan formal level [Pia53]).

Of course, the above considerations are very far from a rigorous mathematical proof (or even precise formulation) of this hypothesis. But we are presenting this here as a conceptual hypothesis, in order to qualitatively guide our practical AGI R&D and also to motivate further, more rigorous theoretical work.

9.3.1 Generalizing the Embodied Communication Prior

One interesting direction for further research would be to broaden the scope of the inquiry, in a manner suggested above: instead of just looking at the ECP, look at simplicity measures in general, and attack the question of how a mind must be structured in order to display efficient general intelligence relative to a specified simplicity measure. This problem seems unapproachable in general, but some special cases may be more tractable.

For instance, suppose one has

- a simplicity measure that (like the ECP) is approximately decomposable into a set of fairly distinct components, plus their interactions
- an assumption similar to NKC, which states that the entities displaying simplicity according to each of the distinct components, are roughly clustered together in entity-space

Then one should be able to say that, to achieve efficient general intelligence relative to this decomposable simplicity measure, a system should have distinct capabilities corresponding to each of the components of the simplicity measure interactions between these capabilities, corresponding to the interaction terms in the simplicity measure.

With copious additional work, these simple observations could potentially serve as the seed for a novel sort of theory of general intelligence – a theory of how the structure of a system depends on the structure of the simplicity measure with which it achieves efficient general intelligence. Cognitive Synergy Theory would then emerge as a special case of this more abstract theory.

9.4 Naive Physics

Multimodal communication is an important aspect of the environment for which human intelligence evolved – but not the only one. It seems likely that our human intelligence is also closely adapted to various aspects of our physical environment – a matter that is worth carefully attending as we design environments for our robotically or virtually embodied AGI systems to operate in.

One interesting guide to the most cognitively relevant aspects of human environments is the subfield of AI known as "naive physics" [Hay85] – a term that refers to the theories about the physical world that human beings implicitly develop and utilize during their lives. For instance, when you figure out that you need to pressure the knife slightly harder when spreading peanut butter rather than jelly, you're not making this judgment using Newtonian physics or the Navier-Stokes equations of fluid dynamics; you're using heuristic patterns that you figured out through experience. Maybe you figured out these patterns through experience spreading peanut butter and jelly in particular. Or maybe you figured these heuristic patterns out before you ever tried to spread peanut butter or jelly specifically, via just touching peanut butter and jelly to see what they feel like, and then carrying out inference based on your experience manipulating similar tools in the context of similar substances.

Other examples of similar "naive physics" patterns are easy to come by, e.g.

1. What goes up must come down.
2. A dropped object falls straight down.
3. A vacuum sucks things towards it.
4. Centrifugal force throws rotating things outwards.
5. An object is either at rest or moving, in an absolute sense.
6. Two events are simultaneous or they are not.
7. When running downhill, one must lift one's knees up high.
8. When looking at something that you just barely can't discern accurately, squint.

Attempts to axiomatically formulate naive physics have historically come up short, and we doubt this is a promising direction for AGI. However, we do think the naive physics literature does a good job of identifying the various phenomena that the human mind's naive physics deals with. So, from the point of view of AGI environment design, naive physics is a useful source of requirements. Ideally, we would like an AGI's environment to support all the fundamental phenomena that naive physics deals with.

We now describe some key aspects of naive physics in a more systematic manner. Naive physics has many different formulations; in this section we draw heavily on [SC94], who divide naive physics phenomena into 5 categories. Here we review these categories and identify a number of important things that humanlike intelligent agents must be able to do relative to each of them.

9.4.1 Objects, Natural Units and Natural Kinds

One key aspect of naive physics involves recognition of various aspects of objects, such as:

1. Recognition of objects amidst noisy perceptual data
2. Recognition of surfaces and interiors of objects
3. Recognition of objects as manipulable units
4. Recognition of objects as potential subjects of fragmentation (splitting, cutting) and of unification (gluing, bonding)
5. Recognition of the agent's body as an object, and as parts of the agent's body as objects
6. Division of universe of perceived objects into "natural kinds", each containing typical and atypical instances

9.4.2 Events, Processes and Causality

Specific aspects of naive physics related to temporality and causality are:

1. Distinguishing roughly-subjectively-instantaneous events from extended processes
2. Identifying beginnings, endings and crossings of processes
3. Identifying and distinguishing internal and external changes
4. Identifying and distinguishing internal and external changes relative to one's own body
5. Interrelating body-changes with changes in external entities

Notably, these aspects of naive physics involve a different processes occurring on a variety of different time scales, intersecting in complex patterns, and involving processes inside the agent's body, outside the agent's body, and crossing the boundary of the agent's body.

9.4.3 Stuffs, States of Matter, Qualities

Regarding the various states of matter, some important aspects of naive physics are:

1. Perceiving gaps between objects: holes, media, illusions like rainbows, mirages and holograms
2. Distinguishing the manners in which different sorts of entities (e.g. smells, sounds, light) fill space
3. Distinguishing properties such as smoothness, roughness, graininess, stickiness, runniness, etc.
4. Distinguishing degrees of elasticity and fragility
5. Assessing separability of aggregates

9.4.4 Surfaces, Limits, Boundaries, Media

Gibson [Gib77, Gib79] has argued that naive physics is not mainly about objects but rather mainly about surfaces. Surfaces have a variety of aspects and relationships that are important for naive physics, such as:

1. Perceiving and reasoning about surfaces as two-sided or one-sided interfaces
2. Inference of the various ecological laws of surfaces
3. Perception of various media in the world as separated by surfaces
4. Recognition of the textures of surfaces
5. Recognition of medium/surface layout relationships such as: ground, open environment, enclosure, detached object, attached object, hollow object, place, sheet, fissure, stick, fibre, dihedral, etc.

As a concrete, evocative "toy" example of naive everyday knowledge about surfaces and boundaries, consider Sloman's [Slo08a] example scenario, depicted in Figure 9.1 and drawn largely from [SS74] (see also related discussion in [Slo08b], in which "A child can be given one

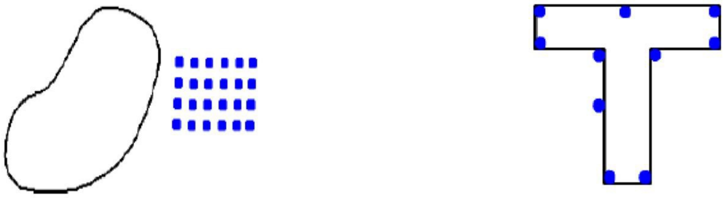


Fig. 9.1: One of Sloman's example test domains for real-world inference. Left: a number of pins and a rubber band to be stretched around them. Right: use of the pins and rubber band to make a letter T.

or more rubber bands and a pile of pins, and asked to use the pins to hold the band in place to form a particular shape)... For example, things to be learnt could include":

1. There is an area inside the band and an area outside the band.
2. The possible effects of moving a pin that is inside the band towards or further away from other pins inside the band. (The effects can depend on whether the band is already stretched.)
3. The possible effects of moving a pin that is outside the band towards or further away from other pins inside the band.
4. The possible effects of adding a new pin, inside or outside the band, with or without pushing the band sideways with the pin first.
5. The possible effects of removing a pin, from a position inside or outside the band.
6. Patterns of motion/change that can occur and how they affect local and global shape (e.g. introducing a concavity or convexity, introducing or removing symmetry, increasing or decreasing the area enclosed).
7. The possibility of causing the band to cross over itself. (NB: Is an odd number of crosses possible?)
8. How adding a second, or third band can enrich the space of structures, processes and effects of processes.

9.4.5 What Kind of Physics Is Needed to Foster Human-like Intelligence?

We stated above that we would like an AGI's environment to support all the fundamental phenomena that naive physics deals with; and we have now reviewed a number of these specific phenomena. But it's not entirely clear what the "fundamental" aspects underlying these phenomena are. One important question in the environment-design context is how close an AGI environment needs to stick to the particulars of real-world naive physics. Is it important that a young AGI can play with the specific differences between spreading peanut butter versus jelly? Or is it enough that it can play with spreading and smearing various substances of different consistencies? How close does the analogy between an AGI environment's naive physics and real-world naive physics need to be? This is a question to which we have no scientific answer at present. Our own working hypothesis is that the analogy does not need to be extremely close, and with this in mind in Chapter 16 we propose a virtual environment BlocksNBeadsWorld that encompasses all the basic conceptual phenomena of real-world naive physics, but does not attempt to emulate their details.

Framed in terms of human psychology rather than environment design, the question becomes: *At what level of detail must one model the physical world to understand the ways in which human intelligence has adapted to the physical world?*. Our suspicion, which underlies our BlocksNBeadsWorld design, is that it's approximately enough to have

- Newtonian physics, or some close approximation
- Matter in multiple phases and forms vaguely similar to the ones we see in the real world: solid, liquid, gas, paste, goo, etc.
- Ability to transform some instances of matter from one form to another
- Ability to flexibly manipulate matter in various forms with various solid tools
- Ability to combine instances of matter into new ones in a fairly rich way: e.g. glue or tie solids togethermix liquids together, etc.
- Ability to position instances of matter with respect to each other in a rich way: e.g. put liquid in a solid cavity, cover something with a lid or a piece of fabric, etc.

It seems to us that if the above are present in an environment, then an AGI seeking to achieve appropriate goals in that environment will be likely to form an appropriate "human-like physical-world intuition." We doubt that the specifics of the naive physics of different forms of matter are critical to human-like intelligence. But, we suspect that a great amount of unconscious human metaphorical thinking is conditioned on the fact that humans evolved around matter that takes a variety of forms, can be changed from one form to another, and can be fairly easily arranged and composited to form new instances from prior ones. Without many diverse instances of matter transformation, arrangement and composition in its experience, an AGI is unlikely to form an internal "metaphor-base" even vaguely similar to the human one – so that, even if it's highly intelligent, its thinking will be radically non-human-like in character.

Naturally this is all somewhat speculative and must be explored via experimentation. Maybe an elaborate blocks-world with only solid objects will be sufficient to create human-level, roughly human-like AGI with rich spatiotemporal and manipulative intuition. Or maybe human intelligence is more closely adapted to the specifics of our physical world – with water and dirt and plants and hair and so forth – than we currently realize. One thing that *is* very clear is that, as we proceed with embodying, situating and educating our AGI systems, we need to pay careful attention to the way their intelligence is conditioned by their environment.

9.5 Folk Psychology

Related to naive physics is the notion of "naive psychology" or "folk psychology" [Rav04], which includes for instance the following aspects:

1. Mental simulation of other agents
2. Mental theory regarding other agents
3. Attribution of beliefs, desires and intentions (BDI) to other agents via theory or simulation

4. Recognition of emotions in other agents via their physical embodiment
5. Recognition of desires and intentions in other agents via their physical embodiment
6. Analogical and contextual inferences between self and other, regarding BDI and other aspects
7. Attribute causes and meanings to other agents behaviors
8. Anthropomorphize non-human, including inanimate objects

The main special requirement placed on an AGI's embodiment by the above aspects pertains to the ability of agents to express their emotions and intentions to each other. Humans do this via facial expressions, gestures and language.

9.5.1 Motivation, Requiredness, Value

Relatedly to folk psychology, Gestalt [Koh38] and ecological [Gib77, Gib79] psychology suggest that humans perceive the world substantially in terms of the affordances it provides them for goal-directed action. This suggests that, to support human-like intelligence, an AGI must be capable of:

1. Perception of entities in the world as differentially associated with goal-relevant value
2. Perception of entities in the world in terms of the potential actions they afford the agent, or other agents

The key point is that entities in the world need to provide a wide variety of ways for agents to interact with them, enabling richly complex perception of affordances.

9.6 Body and Mind

The above discussion has focused on the world external to the body of the AGI agent embodied and embedded in the world, but the issue of the AGI's body also merits consideration. There seems little doubt that a human's intelligence is highly conditioned by the particularities of the human body.

9.6.1 The Human Sensorium

Here the requirements seem fairly simple: while surely not strictly necessary, it would certainly be *preferable* to provide an AGI with fairly rich analogues of the human senses of touch, sight, sound, kinesthesia, taste and smell. Each of these senses provides different sorts of cognitive stimulation to the human mind; and while similar cognitive stimulation could doubtless be achieved without analogous senses, the provision of such seems the most straightforward approach. It's hard to know how much of human intelligence is specifically biased to the sorts of outputs provided by human senses.

As vision already is accorded such a prominent role in the AI and cognitive science literature – and is discussed in moderate depth in Chapter 26 of Part 2, we won't take time elaborating on the importance of vision processing for humanlike cognition. The key thing an AGI requires to support humanlike "visual intelligence" is an environment containing a sufficiently robust collection of materials that object and event recognition and identification become interesting problems.

Audition is cognitively valuable for many reasons, one of which is that it gives a very rich and precise method of sensing the world that is different from vision. The fact that humans can display normal intelligence while totally blind or totally deaf is an indication that, in a sense, vision and audition are redundant for understanding the everyday world. However, it may be important that the brain has evolved to account for both of these senses, because this forced it to account for the presence of two very rich and precise methods of sensing the world – which may have forced it to develop more abstract representation mechanisms than would have been necessary with only one such method.

Touch is a sense that is, in our view, generally badly underappreciated within the AI community. In particular the cognitive robotics community seems to worry too little about the terribly impoverished sense of touch possessed by most current robots (though fortunately there are recent technologies that may help improve robots in this regard; see e.g. [Nan08]). Touch is how the human infant learns to distinguish self from other, and in this way it is the most essential sense for the establishment of an internal self-model. Touching others' bodies is a key method for developing a sense of the emotional reality and responsiveness of others, and is hence key to the development of theory of mind and social understanding in humans. For this reason, among others, human children lacking sufficient tactile stimulation will generally wind up badly impaired in multiple ways. A good-quality embodiment should supply an AI agent with a body that possesses skin, which has varying levels of sensitivity on different parts of the skin (so that it can effectively distinguish between reality and its perception thereof in a tactile context); and also varying types of touch sensors (e.g. temperature versus friction), so that it experiences textures as multidimensional entities.

Related to touch, kinesthesia refers to direct sensation of phenomena happening inside the body. Rarely mentioned in AI, this sense seems quite critical to cognition, as it underpins many of the analogies between self and other that guide cognition. Again, it's not important that an AGI's virtual body have the same internal body parts as a human body. But it seems valuable to have the AGI's virtual body display some vaguely human-body-like properties, such as feeling internal strain of various sorts after getting exercise, feeling discomfort in certain places when running out of energy, feeling internally different when satisfied versus unsatisfied, etc.

Next, taste is a cognitively interesting sense in that it involves the interplay between the internal and external world; it involves the evaluation of which entities from the external world are worthy of placing inside the body. And smell is cognitively interesting in large part because of its relationship with taste. A smell is, among other things, a long-distance indicator of what a certain entity might taste like. So, the combination of taste and smell provides means for conceptualizing relationships between self, world and distance.

9.6.2 The Human Body's Multiple Intelligences

While most unique aspect of human intelligence is rooted in what one might call the "cognitive cortex" – the portions of the brain dealing with self-reflection and abstract thought. But the cognitive cortex does its work in close coordination with the body's various more specialized intelligent subsystems, including those associated with the gut, the heart, the liver, the immune and endocrine systems, and the perceptual and motor cortices.

In the perspective underlying this book, the human cognitive cortex – or the core cognitive network of any roughly human-like AGI system – should be viewed as a highly flexible, self-organizing network. These cognitive networks are modelable e.g. as a recurrent neural net with general topology, or a weighted labeled hypergraph, and are centrally concerned with recognizing patterns in its environment and itself, especially patterns regarding the achievement of the system's goals in various appropriate contexts. Here we augment this perspective, noting that the human brain's cognitive network is closely coupled with a variety of simpler and more specialized intelligent "body-system networks" which provide it with structural and dynamical inductive biasing. We then discuss the implications of this observation for practical AGI design.

One recalls Pascal's famous quote "The heart has its reasons, of which reason knows not." As we now know, the intuitive sense that Pascal and so many others have expressed, that the heart and other body systems have their own reasons, is grounded in the fact that they actually do carry out simple forms of reasoning (i.e. intelligent, adaptive dynamics), in close, sometimes cognitively valuable, coordination with the central cognitive network.

9.6.2.1 Some of the Human Body's Specialized Intelligent Subsystems

The human body contains multiple specialized intelligences apart from the cognitive cortex. Here we review some of the most critical.

Hierarchies of Visual and Auditory Perception

. The hierarchical structure of visual and auditory cortex has been taken by some researchers [Kur12], [HB06] as the generic structure of cognition. While we suspect this is overstated, we agree it is important that these cortices nudge large portions of the cognitive cortex to assume an approximately hierarchical structure.

Olfactory Attractors

. The process of recognizing a familiar smell is grounded in a neural process similar to convergence to an attractor in a nonlinear dynamical system [Fre95]. There is evidence that the mammalian cognitive cortex evolved in close coordination with the olfactory cortex [Row11], and much of abstract cognition reflects a similar dynamic of gradually coming to a conclusion based on what initially "smells right."

Physical and Cognitive Action

. The cerebellum, a specially structured brain subsystem which controls motor movements, has for some time been understood to also have involvement in attention, executive control, language, working memory, learning, pain, emotion, and addiction [PSF09].

The Second Brain

. The gastrointestinal neural net contains millions of neurons and is capable of operating independently of the brain. It modulates stress response and other aspects of emotion and motivation based on experience – resulting in so-called "gut feelings" [Ger99].

The Heart's Neural Network

. The heart has its own neural network, which modulates stress response, energy level and relaxation/excitement (factors key to motivation and emotion) based on experience [Arm04].

Pattern Recognition and Memory in the Liver

. The liver is a complex pattern recognition system, adapting via experience to better identify toxins [CB06]. Like the heart, it seems to store some episodic memories as well, resulting in liver transplant recipients sometimes acquiring the tastes in music or sports of the donor [EMC12].

Immune Intelligence

. The immune network is a highly complex, adaptive self-organizing system, which ongoingly solves the learning problem of identifying antigens and distinguishing them from the body system [FP86]. As immune function is highly energetically costly, stress response involves subtle modulation of the energy allocation to immune function, which involves communication between neural and immune networks.

The Endocrine System: A Key Bridge Between Mind and Body

. The endocrine (hormonal) system regulates (and is related by) emotion, thus guiding all aspects of intelligence (due to the close connection of emotion and motivation) [PH12].

Breathing Guides Thinking

. As oxygenation of the brain plays a key role in the spread of neural activity, the flow of breath is a key driver of cognition. Forced alternate nostril breathing has been shown to significantly affect cognition via balancing activity of the two brain hemispheres [SKBB91].

Much remains unknown, and the totality of feedback loops between the human cognitive cortex and the various specialized intelligences operative throughout the human body, has not yet been thoroughly charted.

9.6.2.2 Implications for AGI

What lesson should the AGI developer draw from all this? The particularities of the human mind/body should not be taken as general requirements for general intelligence. However, it is worth remembering just how difficult is the computational problem of learning, based on experiential feedback alone, the right way to achieve the complex goal of controlling a system with general intelligence at the human level or beyond. To solve this problem without some sort of strong inductive biasing may require massively more experience than young humans obtain.

Appropriate inductive bias may be embedded in an AGI system in many different ways. Some AGI designers have sought to embed it very explicitly, e.g. with hand-coded declarative knowledge as in Cyc, SOAR and other "GOFAI" type systems. On the other hand, the human brain receives its inductive bias much more subtly and implicitly, both via the specifics of the initial structure of the cognitive cortex, and via ongoing coupling of the cognitive cortex with other systems possessing more focused types of intelligence and more specific structures and/or dynamics.

In building an AGI system, one has four choices, very broadly speaking:

1. Create a flexible mind-network, as unbiased as feasible, and attempt to have it learn how to achieve its goals via experience
2. Closely emulate key aspects of the human body along with the human mind
3. Imitate the human mind-body, conceptually if not in detail, and create a number of structurally and dynamically simpler intelligent systems closely and appropriately coupled to the abstract cognitive mind-network, provide useful inductive bias.
4. Find some other, creative way to guide and probabilistically constrain one's AGI system's mind-network, providing inductive bias appropriate to the tasks at hand, without emulating even conceptually the way the human mind-brain receives its inductive bias via coupling with simpler intelligent systems.

Our suspicion is that the first option will not be viable. On the other hand, to do the second option would require more knowledge of the human body than biology currently possesses. This leaves the third and fourth options, both of which seem viable to us.

CogPrime incorporates a combination of the third and fourth options. CogPrime's generic dynamic knowledge store, the Atomspace, is coupled with specialized hierarchical networks (DeSTIN) for vision and audition, somewhat mirroring the human cortex. An artificial endocrine system for OpenCog is also under development, speculatively, as part of a project using OpenCog to control humanoid robots. On the other hand, OpenCog has no gastrointestinal nor cardiological nervous system, and the stress-response-based guidance provided to the human brain by a combination of the heart, gut, immune system and other body systems, is achieved in CogPrime in a more explicit way using the OpenPsi model of motivated cognition, and its integration with the system's attention allocation dynamics.

Likely there is no single correct way to incorporate the lessons of intelligent human body-system networks into AGI designs. But these are aspects of human cognition that all AGI researchers should be aware of.

9.7 The Extended Mind and Body

Finally, Hutchins [Hut95], Logan [Log07] and others have promoted a view of human intelligence that views the human mind as extended beyond the individual body, incorporating social interactions and also interactions with inanimate objects, such as tools, plants and animals. This leads to a number of requirements for a humanlike AGI's environment:

1. The ability to create a variety of different tools for interacting with various aspects of the world in various different ways, including tools for making tools and ultimately machinery
2. The existence of other mobile, virtual life-forms in the world, including simpler and less intelligent ones, and ones that interact with each other and with the AGI
3. The existence of organic growing structures in the world, with which the AGI can interact in various ways, including halting their growth or modifying their growth pattern

How necessary these requirements are is hard to say – but it *is* clear that these things have played a major role in the evolution of human intelligence.

9.8 Conclusion

Happily, this diverse chapter supports a simple, albeit tentative conclusion. Our suggestion is that, if an AGI is

- placed in an environment capable of roughly supporting multimodal communication and vaguely (but not necessarily precisely) real-world-ish naive physics
- surrounded with other intelligent agents of varying levels of complexity, and other complex, dynamic structures to interface with
- given a body that can perceive this environment through some forms of sight, sound and touch; and perceive itself via some form of kinesthesia
- given a motivational system that encourages it to make rich use of these aspects of its environment

then the AGI is likely to have an experience-base reinforcing the key inductive biases provided by the everyday world for the guidance of humanlike intelligence.

Chapter 10
A Mind-World Correspondence Principle

10.1 Introduction

Real-world minds are always adapted to certain classes of environments and goals. The ideas of the previous chapter, regarding the connection between a human-like intelligence's internals and its environment, result from exploring the implications of this adaptation in the context of the cognitive synergy concept. In this chapter we explore the mind-world connection in a broader and more abstract way – making a more ambitious attempt to move toward a "general theory of general intelligence."

One basic premise here, as in the preceding chapters is: Even a system of vast general intelligence, subject to real-world space and time constraints, will necessarily be more efficient at some kinds of learning than others. Thus, one approach to formulating a general theory of general intelligence is to look at the relationship between minds and worlds – where a "world" is conceived as an environment and a set of goals defined in terms of that environment.

In this spirit, we here formulate a broad principle binding together worlds and the minds that are intelligent in these worlds. The ideas of the previous chapter constitute specific, concrete instantiations of this general principle. A careful statement of the principle requires introduction of a number of technical concepts, and will be given later on in the chapter. A crude, informal version of the principle would be:

MIND-WORLD CORRESPONDENCE-PRINCIPLE

For a mind to work intelligently toward certain goals in a certain world, there should be a nice mapping from goal-directed sequences of world-states into sequences of mind-states, where "nice" means that a world-state-sequence W composed of two parts W_1 and W_2, gets mapped into a mind-state-sequence M composed of two corresponding parts M_1 and M_2.

What's nice about this principle is that it relates the decomposition of the world into parts, to the decomposition of the mind into parts.

10.2 What Might a General Theory of General Intelligence Look Like?

It's not clear, at this point, what a real "general theory of general intelligence" would look like – but one tantalizing possibility is that it might confront the two questions:

- How does one design a world to foster the development of a certain sort of mind?
- How does one design a mind to match the particular challenges posed by a certain sort of world?

One way to achieve this would be to create a theory that, given a description of an environment and some associated goals, would output a description of the structure and dynamics that a system should possess to be intelligent in that environment relative to those goals, using limited computational resources.

Such a theory would serve a different purpose from the mathematical theory of "universal intelligence" developed by Marcus Hutter [Hut05] and others. For all its beauty and theoretical power, that approach currently gives it useful conclusions only about general intelligences with infinite or infeasibly massive computational resources. On the other hand, the approach suggested here is aimed toward creation of a theory of real-world general intelligences utilizing realistic amounts of computational power, but still possessing general intelligence comparable to human beings or greater.

This reflects a vision of intelligence as largely concerned with adaptation to particular classes of environments and goals. This may seem contradictory to the notion of "general" intelligence, but I think it actually embodies a realistic understanding of general intelligence. Maximally general intelligence is not pragmatically feasible; it could only be achieved using infinite computational resources [Hut05]. Real-world systems are inevitably limited in the intelligence they can display in any real situation, because real situations involve finite resources, including finite amounts of time. One may say that, in principle, a certain system could solve any problem given enough resources and time but, even when this is true, it's not necessarily the most interesting way to look at the system's intelligence. It may be more important to look at what a system can do given the resources at its disposal in reality. And this perspective leads one to ask questions like the ones posed above: which bounded-resources systems are well-disposed to display intelligence in which classes of situations?

As noted in Chapter 7 above, one can assess the generality of a system's intelligence via looking at the entropy of the class of situations across which it displays a high level of intelligence (where "high" is measured relative to its total level of intelligence across all situations). A system with a high generality of intelligence will tend to be roughly equally intelligent across a wide variety of situations; whereas a system with lower generality of intelligence will tend to be much more intelligent in a small subclass of situations, than in any other. The definitions given above embody this notion in a formal and quantitative way.

If one wishes to create a general theory of general intelligence according to this sort of perspective, the main question then becomes how to represent goals/environments and systems in such a way as to render transparent the natural correspondence between the specifics of the former and the latter, in the context of resource-bounded intelligence. This is the business of the next section.

10.3 Steps Toward A (Formal) General Theory of General Intelligence

Now begins the formalism. At this stage of development of the theory proposed in this chapter, mathematics is used mainly as a device to ensure clarity of expression. However, once the theory is further developed, it may possibly become useful for purposes of calculation as well.

Suppose one has any system S (which could be an AI system, or a human, or an environment that a human or AI is interacting with, or the combination of an environment and a human or AI's body, etc.). One may then construct an uncertain transition graph associated with that system S, in the following way:

- The nodes of the graph represent fuzzy sets of states of system S (I'll call these state-sets from here on, leaving the fuzziness implicit)
- The (directed) links of the graph represent probabilistically weighted transitions between state-sets

Specifically, the weight of the link from B to A should be defined as

$$P(o(S,A,t(T))|o(S,B,T))$$

where

$$o(S,A,T)$$

denotes the presence of the system S in the state-set A during time-distribution T, and $t()$ is a temporal succession function defined so that $t(T)$ refers to a time-distribution conceived as "after" T. A time-distribution is a probability distribution over time-points. The interaction of fuzziness and probability here is fairly straightforward and may be handled in the manner of PLN, as outlined in subsequent chapters. Note that the definition of link weights is dependent on the specific implementation of the temporal succession function, which includes an implicit time-scale.

Suppose one has a transition graph corresponding to an environment; then a goal relative to that environment may be defined as a particular node in the transition graph. The goals of a particular system acting in that environment may then be conceived as one or more nodes in the transition graph. The system's situation in the environment at any point in time may also be associated with one or more nodes in the transition graph; then, the system's movement toward goal-achievement may be associated with paths through the environment's transition graph leading from its current state to goal states.

It may be useful for some purposes to filter the uncertain transition graph into a crisp transition graph by placing a threshold on the link weights, and removing links with weights below the threshold.

The next concept to introduce is the world-mind transfer function, which maps world (environment) state-sets into organism (e.g. AI system) state-sets in a specific way. Given a world state-set W, the world-mind transfer function M maps W into various organism state-sets with various probabilities, so that we may say: $M(W)$ is the probability distribution of state-sets the organism tends to be in, when its environment is in state-set W. (Recall also that state-sets are fuzzy.)

Now one may look at the spaces of world-paths and mind-paths. A world-path is a path through the world's transition graph, and a mind-path is a path through the organism's transition graph. Given two world-paths P and Q, it's obvious how to define the composition $P*Q$ one follows P and then, after that, follows Q, thus obtaining a longer path. Similarly for mind-paths.

In category theory terms, we are constructing the free category associated with the graph: the objects of the category are the nodes, and the morphisms of the category are the paths. And category theory is the right way to be thinking here we want to be thinking about the relationship between the world category and the mind category.

The world-mind transfer function can be interpreted as a mapping from paths to subgraphs: Given a world-path, it produces a set of mind state-sets, which have a number of links between them. One can then define a world-mind path transfer function $M(P)$ via taking the mind-graph $M(nodes(P))$, and looking at the highest-weight path spanning $M(nodes(P))$. (Here $nodes?$ obviously means the set of nodes of the path P.)

A functor F between the world category and the mind category is a mapping that preserves object identities and so that

$$F(P * Q) = F(P) * F(Q)$$

We may also introduce the notion of an approximate functor, meaning a mapping F so that the average of

$$d(F(P * Q), F(P) * F(Q))$$

is small.

One can introduce a prior distribution into the average here. This could be the Levin universal distribution or some variant (the Levin distribution assigns higher probability to computationally simpler entities). Or it could be something more purpose specific: for example, one can give a higher weight to paths leading toward a certain set of nodes (e.g. goal nodes). Or one can use a distribution that weights based on a combination of simplicity and directedness toward a certain set of nodes. The latter seems most interesting, and I will define a goal-weighted approximate functor as an approximate functor, defined with averaging relative to a distribution that balances simplicity with directedness toward a certain set of goal nodes.

The move to approximate functors is simple conceptually, but mathematically it's a fairly big step, because it requires us to introduce a geometric structure on our categories. But there are plenty of natural metrics defined on paths in graphs (weighted or not), so there's no real problem here.

10.4 The Mind-World Correspondence Principle

Now we finally have the formalism set up to make a non-trivial statement about the relationship between minds and worlds. Namely, the hypothesis that:

MIND-WORLD CORRESPONDENCE PRINCIPLE

For an organism with a reasonably high level of intelligence in a certain world, relative to a certain set of goals, the mind-world path transfer function is a goal-weighted approximate functor.

That is, a little more loosely: the hypothesis is that, for intelligence to occur, there has to be a natural correspondence between the transition-sequences of world-states and the corresponding transition-sequences of mind-states, at least in the cases of transition-sequences leading to relevant goals.

We suspect that a variant of the above proposition can be formally proved, using the definition of general intelligence presented in Chapter 7. The proof of a theorem corresponding to the above would certainly constitute an interesting start toward a general formal theory of general intelligence. Note that proving anything of this nature would require some attention to the time-scale-dependence of the link weights in the transition graphs involved.

A formally proved variant of the above proposition would be in short, a "MIND-WORLD CORRESPONDENCE THEOREM."

Recall that at the start of the chapter, we expressed the same idea as:

MIND-WORLD CORRESPONDENCE-PRINCIPLE

For a mind to work intelligently toward certain goals in a certain world, there should be a nice mapping from goal-directed sequences of world-states into sequences of mind-states, where "nice" means that a world-state-sequence W composed of two parts W_1 and W_2, gets mapped into a mind-state-sequence M composed of two corresponding parts M_1 and M_2.

That is a reasonable gloss of the principle, but it's clunkier and less accurate, than the statement in terms of functors and path transfer functions, because it tries to use only common-language vocabulary, which doesn't really contain all the needed concepts.

10.5 How Might the Mind-World Correspondence Principle Be Useful?

Suppose one believes the Mind-World Correspondence Principle as laid out above so what?

Our hope, obviously, is that the principle could be useful in actually figuring out how to architect intelligent systems biased toward particular sorts of environment. And of course, this is said with the understanding that any finite intelligence must be biased toward some sorts of environment.

Relatedly, given a specific AGI design (such as CogPrime), one could use the principle to figure out which environments it would be best suited for. Or one could figure out how to adjust the particulars of the design, to maximize the system's intelligence in the environments of interest.

One next step in developing this network of ideas, aside from (and potentially building on) full formalization of the principle, would be an exploration of real-world environments in terms of transition graphs. What properties do the transition graphs induced from the real world have?

One such property, we suggest, is successive refinement. Often the path toward a goal involves first gaining an approximate understanding of a situation, then a slightly more accurate understanding, and so forth – until finally one has achieved a detailed enough understanding to actually achieve the goal. This would be represented by a world-path whose nodes are state-sets involving the gathering of progressively more detailed information.

Via pursuing to the mind-world correspondence property in this context, I believe we will find that world-paths reflecting successive refinement correspond to mind-paths embodying successive refinement. This will be found to relate to the hierarchical structures found so frequently in both the physical world and the human mind-brain. Hierarchical structures allow many relevant goals to be approached via successive refinement, which I believe is the ultimate reason why hierarchical structures are so common in the human mind-brain.

Another next step would be exploring what mind-world correspondence means for the structure and dynamics of a limited-resources intelligence. If an organism O has limited resources and, to be intelligent, needs to make

$$P(o(O, M(A), t(T))|o(O, M(B), T))$$

high for particular world state-sets A and B, then what's the organism's best approach? Arguably, it should represent $M(A)$ and $M(B)$ internally in such a way that very little computational effort is required for it to transition between $M(A)$ and $M(B)$. For instance, this could be done by coding its knowledge in such a way that $M(A)$ and $M(B)$ share many common bits; or it could be done in other more complicated ways.

If, for instance, A is a subset of B, then it may prove beneficial for the organism to represent $M(A)$ physically as a subset of its representation of $M(B)$.

Pursuing this line of thinking, one could likely derive specific properties of an intelligent organism's internal information-flow, from properties of the environment and goals with respect to which it's supposed to be intelligent.

This would allow us to achieve the holy grail of intelligence theory as I understand it: given a description of an environment and goals, to be able to derive an architectural description for an organism that will display a high level of intelligence relative to those goals, given limited computational resources.

While this "holy grail" is obviously a far way off, what we've tried to do here is to outline a clear mathematical and conceptual direction for moving toward it.

10.6 Conclusion

The Mind-World Correspondence Principle presented here – if in the vicinity of correctness – constitutes a non-trivial step toward fleshing out the concept of a general theory of general intelligence. But obviously the theory is still rather abstract, and also not completely rigorous. There's a lot more work to be done.

The Mind-World Correspondence Principle as articulated above is not quite a formal mathematical statement. It would take a little work to put in all the needed quantifiers to formulate it as one, and it's not clear the best way to do so the details would perhaps become clear in the course of trying to prove a version of it rigorously. One could interpret the ideas presented in this chapter as a philosophical theory that hopes to be turned into a mathematical theory and to play a key role in a scientific theory.

For the time being, the main role to be served by these ideas is qualitative: to help us think about concrete AGI designs like CogPrime in a sensible way. It's important to understand what the goal of a real-world AGI system needs to be: to achieve the ability to broadly learn and generalize, yes, but not with infinite capability rather with biases and patterns that are implicitly and/or explicitly tuned to certain broad classes of goals and environments. The Mind-World

Correspondence Principle tells us something about what this "tuning" should involve – namely, making a system possessing mind-state sequences that correspond meaningfully to world-state sequences. CogPrime's overall design and particular cognitive processes are reasonably well interpreted as an attempt to achieve this for everyday human goals and environments.

One way of extending these theoretical ideas into a more rigorous theory is explored in Appendix **??**. The key ideas involved there are: modeling multiple memory types as mathematical categories (with functors mapping between them), modeling memory items as probability distributions, and measuring distance between memory items using two metrics, one based on algorithmic information theory and one on classical information geometry. Building on these ideas, core hypotheses are then presented:

- a **syntax-semantics correlation** principle, stating that in a successful AGI system, these two metrics should be roughly correlated
- a **cognitive geometrodynamics** principle, stating that on the whole intelligent minds tend to follow geodesics (shortest paths) in mindspace, according to various appropriately defined metrics (e.g. the metric measuring the distance between two entities in terms of the length and/or runtime of the shortest programs computing one from the other).
- a **cognitive synergy** principle, stating that shorter paths may be found through the composite mindspace formed by considering multiple memory types together, than by following the geodesics in the mindspaces corresponding to individual memory types.

The material is relegated to an appendix because it is so speculative, and it's not yet clear whether it will really be useful in advancing or interpreting CogPrime or other AGI systems (unlike the material from the present chapter, which has at least been useful in interpreting and tweaking the CogPrime design, even though it can't be claimed that CogPrime was derived directly from these theoretical ideas). However, this sort of speculative exploration is, in our view, exactly the sort of thing that's needed as a first phase in transitioning the ideas of the present chapter into a more powerful and directly actionable theory.

Section III

Cognitive and Ethical Development

Chapter 11
Stages of Cognitive Development

Co-authored with Stephan Vladimir Bugaj

11.1 Introduction

Creating AGI, we have said, is not only about having the right structural and dynamical possibilities implemented in the initial version of one's system – but also about the environment and embodiment that one's system is associated with, and the match between the system's internals and these externals. Another key aspect is the long-term time-course of the system's evolution over time, both in its internals and its external interaction – i.e., what is known as *development*.

Development is a critical topic in our approach to AGI because we believe that much of what constitutes human-level, human-like intelligence *emerges* in an intelligent system due to its engagement with its environment and its environment-coupled self-organization. So, it's not to be expected that the initial version of an AGI system is going to display impressive feats of intelligence, even if the engineering is totally done right. A good analogy is the apparent unintelligence of a human baby. Yes, scientists have discovered that human babies are capable of interesting and significant intelligence – but one has to hunt to find it ... at first observation, babies are rather idiotic and simple-minded creatures: much less intelligent-appearing than lizards or fish, maybe even less than cockroaches....

If the goal of an AGI project is to create an AGI system that can progressively develop advanced intelligence through learning in an environment richly populated with other agents and various inanimate stimuli and interactive entities – then an understanding of the nature of cognitive development becomes extremely important to that project.

Unfortunately, contemporary cognitive science contains essentially no theory of "abstract developmental psychology" which can conveniently be applied to understand developing AIs. There is of course an extensive science of **human** developmental psychology, and so it is a natural research program to take the chief ideas from the former and inasmuch as possible port them to the AGI domain. This is not an entirely simple matter both because of the differences between humans and AIs and because of the unsettled nature of contemporary developmental psychology theory. But it's a job that must (and will) be done, and the ideas in this chapter may contribute toward this effort.

We will begin here with Piaget's well-known theory of human cognitive development, presenting it in a general systems theory context, then introducing some modifications and extensions and discussing some other relevant work.

11.2 Piagetan Stages in the Context of a General Systems Theory of Development

Our review of AGI architectures in Chapter 4 focused heavily on the concept of **symbolism**, and the different ways in which different classes of cognitive architecture handle symbol representation and manipulation. We also feel that symbolism is critical to the notion of AGI development – and even more broadly, to the systems theory of development in general.

As a broad conceptual perspective on development, we suggest that one may view the development of a complex information processing system, embedded in an environment, in terms of the stages:

- **automatic**: the system interacts with the environment by "instinct", according to its innate programming
- **adaptive**: the system internally adapts to the environment, then interacting with the environment in a more appropriate way
- **symbolic**: the system creates internal symbolic representations of itself and the environment, which in the case of a complex, appropriately structured environment, allows it to interact with the environment more intelligently
- **reflexive**: the system creates internal symbolic representations of its own internal symbolic representations, thus achieving an even higher degree of intelligence

Sketched so broadly, these are not precisely defined categories but rather heuristic, intuitive categories. Formalizing them would be possible but would lead us too far astray here.

One can interpret these stages in a variety of different contexts. Here our focus is the cognitive development of humans and human-like AGI systems, but in Table 11.1 we present them in a slightly more general context, using two examples: the Piagetan example of the human (or humanlike) mind as it develops from infancy to maturity; and also the example of the "origin of life" and the development of life from proto-life up into its modern form. In any event, we allude to this more general perspective on development here mainly to indicate our view that the Piagetan perspective is not something ad hoc and arbitrary, but rather can plausibly be seen as a specific manifestation of more fundamental principles of complex systems development.

11.3 Piaget's Theory of Cognitive Development

The ghost of Jean Piaget hangs over modern developmental psychology in a yet unresolved way. Piaget's theories provide a cogent overarching perspective on human cognitive development, coordinating broad theoretical ideas and diverse experimental results into a unified whole [Pia55]. Modern experimental work has shown Piaget's ideas to be often oversimplified and incorrect. However, what has replaced the Piagetan understanding is not an alternative unified and coherent theory, but a variety of microtheories addressing particular aspects of cognitive development. For this reason a number of contemporary theorists taking a computer science [Shu03] or dynamical systems [Wit07] approach to developmental psychology have chosen to adopt the Piagetan framework in spite of its demonstrated shortcomings, both because of its conceptual strengths and for lack of a coherent, more rigorously grounded alternative.

Our own position is that the Piagetan view of development has some fundamental truth to it, which is reflected via how nicely it fits with a broader view of development in complex systems.

Stage	**General Description**	**Cognitive Development**	**Origin of Life**
Automatic	System-environment information exchange controlled mainly by innate system structures or environment	Piagetan infantile stage	Self-organizing protolife system, e.g. Oparin [Opa52] water droplet, or Cairns-Smith [CS90] clay-based protolife
Adaptive	System-environment info exchange heavily guided by adaptively internally-created system structures	Piagetan "concrete operational" stage: systematic internal world-model guides world-exploration	Simple autopoietic system, e.g. Oparin water droplet w/ basic metabolism
Symbolic	Internal symbolic representation of information exchange process	Piagetan formal stage: explicit logical/experimental learning about how to cognize in various contexts	Genetic code: internal entities that "stand for" aspects of organism and environment, thus enabling complex epigenesis
Reflexive	Thoroughgoing self-modification based on this symbolic representation	Piagetan post-formal stage: purposive self-modification of basic mental processes	Genes+memes: genetic code-patterns guide their own modification via influencing culture

Table 11.1: General Systems Theory of Development: Parallels Between Development of Mind and Origin of Life

Indeed, Piaget viewed developmental stages as emerging from general "algebraic" principles rather than as being artifacts of the particulars of human psychology. But, Piaget's stages are probably best viewed as a general interpretive framework rather than a precise scientific theory. Our suspicion is that once the empirical science of developmental psychology has progressed further, it will become clearer how to fit the various data into a broad Piaget-like framework, perhaps differing in many details from what Piaget described in his works.

Piaget conceived of child development in four stages, each roughly identified with an age group, and corresponding closely to the system-theoretic stages mentioned above:

- **infantile**, corresponding to the automatic stage mentioned above
  - *Example:* Grasping blocks, piling blocks on top of each other, copying words that are heard
- **preoperational** and **concrete operational**, corresponding to the adaptive stage mentioned above
  - *Example:* Building complex blocks structures, from imagination and from imitating objects and pictures and based on verbal instructions; verbally describing what has been constructed
- **formal**, corresponding to the symbolic stage mentioned above
  - *Example:* Writing detailed instructions in words and diagrams, explaining how to construct particular structures out of blocks; figuring out general rules describing which sorts of blocks structures are likely to be most stable

- the reflexive stage mentioned above corresponds to what some post-Piagetan theorists have called the **post-formal** stage
    - *Example:* Using abstract lessons learned from building structures out of blocks to guide the construction of new ways to think and understand – "Zen and the art of blocks building" (by analogy to *Zen and the Art of Motorcycle Maintenance* [Pir84]).

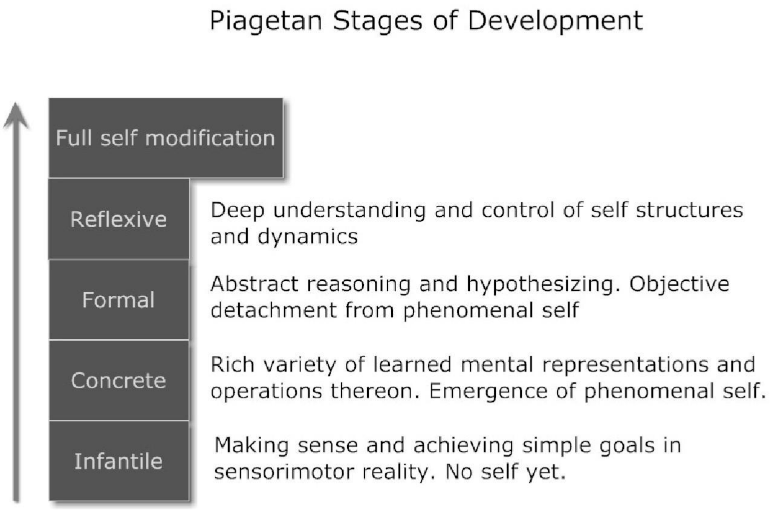

Piagetan Stages of Development
Full self modification
Reflexive
Deep understanding and control of self structures and dynamics
Formal
Abstract reasoning and hypothesizing. Objective detachment from phenomenal self
Concrete
Rich variety of learned mental representations and operations thereon. Emergence of phenomenal self.
Infantile
Making sense and achieving simple goals in sensorimotor reality. No self yet.

Fig. 11.1: Piagetan Stages of Cognitive Development

More explicitly, Piaget defined his stages in psychological terms roughly as follows:

- **Infantile**: In this stage a mind develops basic world-exploration driven by instinctive actions. Reward-driven reinforcement of actions learned by imitation, simple associations between words and objects, actions and images, and the basic notions of time, space, and causality are developed. The most simple, practical ideas and strategies for action are learned.
- **Preoperational**: At this stage we see the formation of mental representations, mostly poorly organized and un-abstracted, building mainly on intuitive rather than logical thinking. Word-object and image-object associations become systematic rather than occasional. Simple syntax is mastered, including an understanding of subject-argument relationships. One of the crucial learning achievements here is "object permanence" – infants learn that objects persist even when not observed. However, a number of cognitive failings persist with respect to reasoning about logical operations, and abstracting the effects of intuitive actions to an abstract theory of operations.
- **Concrete**: More abstract logical thought is applied to the physical world at this stage. Among the feats achieved here are: reversibility – the ability to undo steps already done; conservation – understanding that properties can persist in spite of appearances; theory of mind – an understanding of the distinction between what I know and what others know (If

I cover my eyes, can you still see me?). Complex concrete operations, such as putting items in height order, are easily achievable. Classification becomes more sophisticated, yet the mind still cannot master purely logical operations based on abstract logical representations of the observational world.

- **Formal**: Abstract deductive reasoning, the process of forming, then testing hypotheses, and systematically reevaluating and refining solutions, develops at this stage, as does the ability to reason about purely abstract concepts without reference to concrete physical objects. This is adult human-level intelligence. Note that the capability for formal operations is intrinsic in the PLN component of CogPrime, but in-principle capability is not the same as pragmatic, grounded, controllable capability.

Very early on, Vygotsky [Vyg86] disagreed with Piaget's explanation of his stages as inherent and developed by the child's own activities, and Piaget's prescription of good parenting as not interfering with a child's unfettered exploration of the world. Some modern theorists have critiqued Piaget's stages as being insufficiently socially grounded, and these criticisms trace back to Vygotsky's focus on the social foundations of intelligence, on the fact that children function in a world surrounded by adults who provide a cultural context, offering ongoing assistance, critique, and ultimately validation of the child's developmental activities.

Vygotsky also was an early critic of the idea that cognitive development is continuous, and continues beyond Piaget's formal stage. Gagne [RBW92] also believes in continuity, and that learning of prerequisite skills made the learning of subsequent skills easier and faster without regard to Piagetan stage formalisms. Subsequent researchers have argued that Piaget has merely constructed ad hoc descriptions of the sequential development of behaviour [Gib78, Bro84, CP05]. We agree that learning is a continuous process, and our notion of stages is more statistically constructed than rigidly quantized.

Critique of Piaget's notion of transitional "half stages" is also relevant to a more comprehensive hierarchical view of development. Some have proposed that Piaget's half stages are actually stages [Bro84]. As Commons and Pekker [CP05] point out: "the definition of a stage that was being used by Piaget was based on analyzing behaviors and attempting to impose different structures on them. There is no underlying logical or mathematical definition to help in this process ..." Their Hierarchical Complexity development model uses task achievement rather than ad hoc stage definition as the basis for constructing relationships between phases of developmental ability – an approach which we find useful, though our approach is different in that we define stages in terms of specific underlying cognitive mechanisms.

Another critique of Piaget is that one individual's performance is often at different ability stages depending on the specific task (for example [GE86]). Piaget responded to early critiques along these lines by calling the phenomenon "horizontal décalage," but neither he nor his successors [Fis80, Cas85] have modified his theory to explain (rather than merely describe) it. Similarly to Thelen and Smith [TS94], we observe that the abilities encapsulated in the definition of a certain stage emerge gradually during the previous stage – so that the onset of a given stage represents the mastery of a cognitive skill that was previously present only in certain contexts.

Piaget also had difficulty accepting the idea of a preheuristic stage, early in the infantile period, in which simple trial-and-error learning occurs without significant heuristic guidance [Bic88], a stage which we suspect exists and allows formulation of heuristics by aggregation of learning from preheuristic pattern mining. Coupled with his belief that a mind's innate abilities at birth are extremely limited, there is a troublingly unexplained transition from inability to ability in his model.

Finally, another limiting aspect of Piaget's model is that it did not recognize any stages beyond formal operations, and included no provisions for exploring this possibility. A number of researchers [Bic88, Arl75, CRK82, Rie73, Mar01] have described one or more postformal stages. Commons and colleagues have also proposed a task-based model which provides a framework for explaining stage discrepancies across tasks and for generating new stages based on classification of observed logical behaviors. [KK90] promotes a statistical conception of stage, which provides a good bridge between task-based and stage-based models of development, as statistical modeling allows for stages to be roughly defined and analyzed based on collections of task behaviors.

[CRK82] postulates the existence of a postformal stage by observing *elevated levels of abstraction* which, they argue, are not manifested in formal thought. [CTS+98] observes a postformal stage when subjects become capable of analyzing and coordinating complex logical systems with each other, creating metatheoretical supersystems. In our model, with the reflexive stage of development, we expand this definition of metasystemic thinking to include the ability to consciously refine one's own mental states and formalisms of thinking. Such self-reflexive refinement is necessary for learning which would allow a mind to analytically devise entirely new structures and methodologies for both formal and postformal thinking.

In spite of these various critiques and limitations, however, we have found Piaget's ideas very useful, and in Section 11.4 we will explore ways of defining them rigorously in the specific context of CogPrime's declarative knowledge store and probabilistic logic engine.

11.3.1 Perry's Stages

Also relevant is William Perry's [Per70, Per81] theory of the stages ("positions" in his terminology) of intellectual and ethical development, which constitutes a model of iterative refinement of approach in the developmental process of coming to intellectual and ethical maturity. These stages, depicted in Table 11.2 form an analytical tool for discerning the modality of belief of an intelligence by describing common cognitive approaches to handling the complexities of real world ethical considerations.

11.3.2 Keeping Continuity in Mind

Continuity of mental stages, and the fact that a mind may appear to be in multiple stages of development simultaneously (depending upon the tasks being tested), are crucial to our theoretical formulations and we will touch upon them again here. Piaget attempted to address continuity with the creation of transitional "half stages". We prefer to observe that each stage feeds into the other and the end of one stage and the beginning of the next blend together.

The distinction between formal and post-formal, for example, seems to "merely" be the application of formal thought to oneself. However, the distinction between concrete and formal is "merely" the buildup to higher levels of complexity of the classification, task decomposition, and abstraction capabilities of the concrete stage. The stages represent general trends in ability on a continuous curve of development, not discrete states of mind which are jumped-into quantum style after enough "knowledge energy" builds-up to cause the transition.

Stage	**Substages**
Dualism / Received Knowledge [Infantile]	Basic duality ("All problems are solvable. I must learn the correct solutions.") Full dualism ("There are different, contradictory solutions to many problems. I must learn the correct solutions, and ignore the incorrect ones")
Multiplicity [Concrete]	Early multiplicity ("Some solutions are known, others aren't. I must learn how to find correct solutions.") Late Multiplicity: cognitive dissonance regarding truth. ("Some problems are unsolvable, some are a matter of personal taste, therefore I must declare my own intellectual path.")
Relativism / Procedural Knowledge [Formal]	Contextual Relativism ("I must learn to evaluate solutions within a context, and relative to supporting observation.") Pre-Commitment ("I must evaluate solutions, then commit to a choice of solution.")
Commitment / Constructed Knowledge [Formal / Reflexive]	Commitment ("I have chosen a solution.") Challenges to Commitment ("I have seen unexpected implications of my commitment, and the responsibility I must take.") Post-Commitment ("I must have an ongoing, nuanced relationship to the subject in which I evaluate each situation on a case-by-case basis with respects to its particulars rather than an ad-hoc application of unchallenged ideology.")

Table 11.2: Perry's Developmental Stages [with corresponding Piagetan Stages in brackets]

Observationally, this appears to be the case in humans. People learn things gradually, and show a continuous development in ability, not a quick jump from ignorance to mastery. We believe that this gradual development of ability is the signature of genuine learning, and that prescriptively an AGI system must be designed in order to have continuous and asymmetrical development across a variety of tasks in order to be considered a genuine learning system. While quantum leaps in ability may be possible in an AGI system which can just "graft" new parts of brain onto itself (or an augmented human which may someday be able to do the same using implants), such acquisition of knowledge is not really learning. Grafting on knowledge does not build the cognitive pathways needed in order to actually learn. If this is the only mechanism available to an AGI system to acquire new knowledge, then it is not really a learning system.

11.4 Piaget's Stages in the Context of Uncertain Inference

Piaget's developmental stages are very general, referring to overall types of learning, not specific mechanisms or methods. This focus was natural since the context of his work was *human* developmental psychology, and neuroscience has not yet progressed to the point of understanding the neural mechanisms underlying any sort of inference (and certainly was nowhere near to doing so in Piaget's time!). But if one is studying developmental psychology in an AGI context where one knows something about the internal mechanisms of the AGI system under consideration, then one can work with a more specific model of learning. Our focus here is on AGI systems whose operations contain uncertain inference as a central component. Obviously the main focus is CogPrime, but the essential ideas apply to any other uncertain inference centric AGI architecture as well.

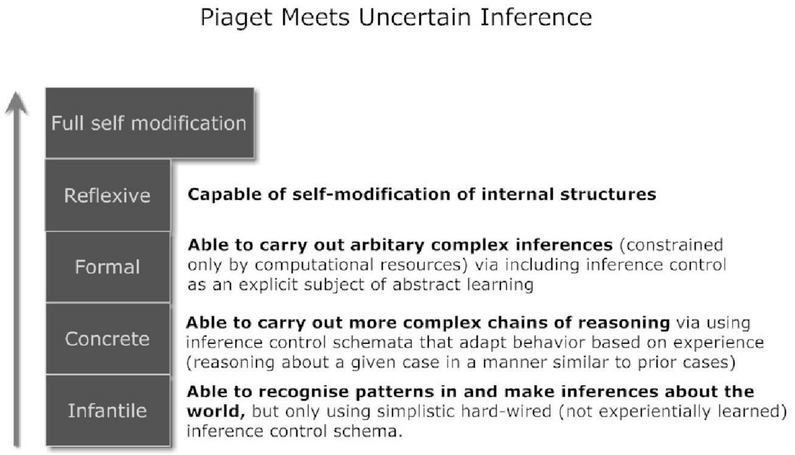

Piaget Meets Uncertain Inference
Full self modification
Reflexive
Capable of self-modification of internal structures
Formal
Able to carry out arbitary complex inferences (constrained only by computational resources) via including inference control as an explicit subject of abstract learning
Concrete
Able to carry out more complex chains of reasoning via using inference control schemata that adapt behavior based on experience (reasoning about a given case in a manner similar to prior cases)
Infantile
Able to recognise patterns in and make inferences about the world, but only using simplistic hard-wired (not experientially learned) inference control schema.

Fig. 11.2: Piagetan Stages of Development, as Manifested in the Context of Uncertain Inference

An uncertain inference system, as we consider it here, consists of four components, which work together in a feedback-control loop 11.3

1. a content representation scheme
2. an uncertainty representation scheme
3. a set of inference rules
4. a set of inference control schemata

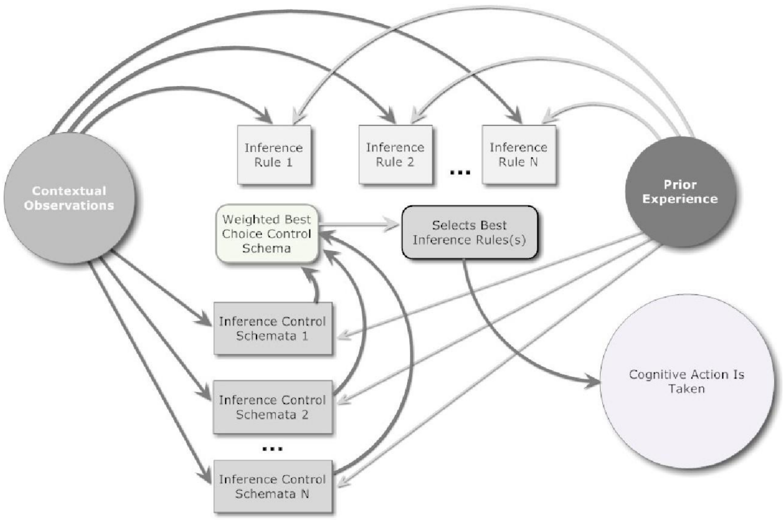

Contextual Observations
Inference Rule 1
Inference Rule 2
...
Inference Rule N
Prior Experience
Weighted Best Choice Control Schema
Selects Best Inference Rules(s)
Inference Control Schemata 1
Inference Control Schemata 2
...
Inference Control Schemata N
Cognitive Action Is Taken

Fig. 11.3: A Simplified Look at Feedback-Control in Uncertain Inference

Broadly speaking, examples of content representation schemes are predicate logic and term logic [ES00]. Examples of uncertainty representation schemes are fuzzy logic [Zad78], imprecise probability theory [Goo86, FC86], Dempster-Shafer theory [Sha76, Kyb97], Bayesian probability theory [Kyb97], NARS [Wan95], and the Atom representation used in CogPrime, briefly alluded to in Chapter 6 above and described in depth in later chapters.

Many, but not all, approaches to uncertain inference involve only a limited, weak set of inference rules (e.g. not dealing with complex quantified expressions). CogPrime's PLN inference framework, like NARS and some other uncertain inference frameworks, contains uncertain inference rules that apply to logical constructs of arbitrary complexity. Only a system capable of dealing with constructs of arbitrary (or at least very high) complexity will have any potential of leading to human-level, human-like intelligence.

The subtlest part of uncertain inference is inference control: the choice of which inferences to do, in what order. Inference control is the primary area in which human inference currently exceeds automated inference. Humans are not very efficient or accurate at carrying out inference rules, with or without uncertainty, but we are very good at determining which inferences to do and in what order, in any given context. The lack of effective, context-sensitive inference control heuristics is why the general ability of current automated theorem provers is considerably weaker than that of a mediocre university mathematics major [Mac95].

We now review the Piagetan developmental stages from the perspective of AGI systems heavily based on uncertain inference.

11.4.1 The Infantile Stage

In this initial stage, the mind is able to recognize patterns in and conduct inferences about the world, but only using simplistic hard-wired (not experientially learned) inference control schema, along with pre-heuristic pattern mining of experiential data.

In the infantile stage an entity is able to recognize patterns in and conduct inferences about its sensory surround context (i.e., it's "world"), but only using simplistic, hard-wired (not experientially learned) inference control schemata. Preheuristic pattern-mining of experiential data is performed in order to build future heuristics about analysis of and interaction with the world.

s tasks include:

1. Exploratory behavior in which useful and useless / dangerous behavior is differentiated by both trial and error observation, and by parental guidance.
2. Development of "habits" – i.e. Repeating tasks which were successful once to determine if they always / usually are so.
3. Simple goal-oriented behavior such as "find out what cat hair tastes like" in which one must plan and take several sequentially dependent steps in order to achieve the goal.

Inference control is very simple during the infantile stage (Figure 11.4), as it is the stage during which both the most basic knowledge of the world is acquired, and the most basic of cognition and inference control structures are developed as the building block upon which will be built the next stages of both knowledge and inference control.

Another example of a cognitive task at the borderline between infantile and concrete cognition is learning object permanence, a problem discussed in the context of CogPrime's predecessor "Novamente Cognition Engine" system in [GPSL03]. Another example is the learning of

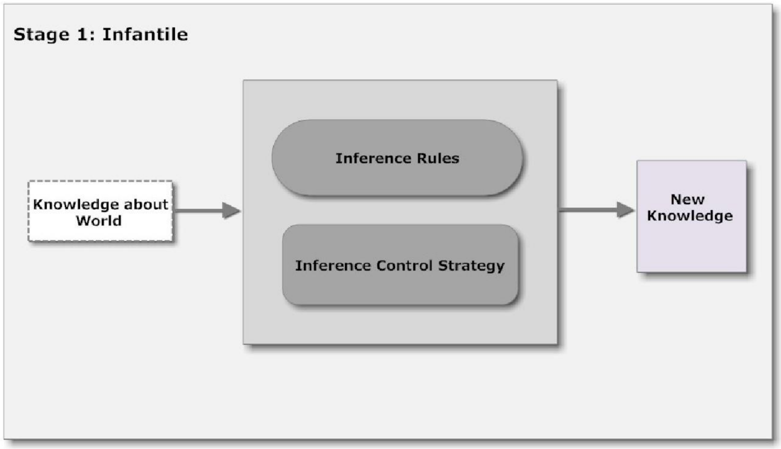


Fig. 11.4: Uncertain Inference in the Infantile Stage

word-object associations: e.g. learning that when the word "ball" is uttered in various contexts ("Get me the ball," "That's a nice ball," etc.) it generally refers to a certain type of object. The key point regarding these "infantile" inference problems, from the CogPrime perspective, is that assuming one provides the inference system with an appropriate set of perceptual and motor ConceptNodes and SchemaNodes, the chains of inference involved are short. They involve about a dozen inferences, and this means that the search tree of possible PLN inference rules walked by the PLN backward-chainer is relatively shallow. Sophisticated inference control is not required: standard AI heuristics are sufficient.

In short, textbook narrow-AI reasoning methods, utilized with appropriate uncertainty-savvy truth value formulas and coupled with appropriate representations of perceptual and motor inputs and outputs, correspond roughly to Piaget's infantile stage of cognition. The simplistic approach of these narrow-AI methods may be viewed as a method of creating building blocks for subsequent, more sophisticated heuristics.

In our theory Piaget's preoperational phase appears as transitional between the infantile and concrete operational phases.

11.4.2 The Concrete Stage

At this stage, the mind is able to carry out more complex chains of reasoning regarding the world, via using inference control schemata that adapt behavior based on experience (reasoning about a given case in a manner similar to prior cases).

In the concrete operational stage (Figure 11.5), an entity is able to carry out more complex chains of reasoning about the world. Inference control schemata which adapt behavior based on experience, using experientially learned heuristics (including those learned in the prior stage), are applied to both analysis of and interaction with the sensory surround / world.

Concrete Operational stage tasks include:

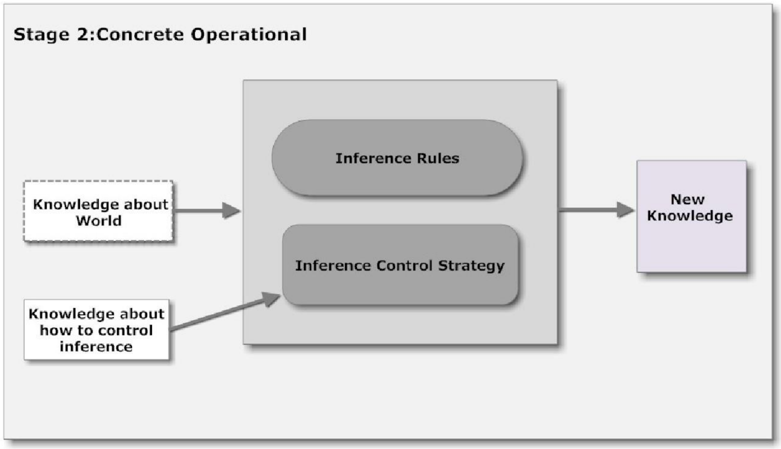

Stage 2:Concrete Operational
Inference Rules
Knowledge about World
Inference Control Strategy
Knowledge about how to control inference
New Knowledge

Fig. 11.5: Uncertain Inference in the Concrete Operational Stage

1. Conservation tasks, such as conservation of number,
2. Decomposition of complex tasks into easier subtasks, allowing increasingly complex tasks to be approached by association with more easily understood (and previously experienced) smaller tasks,
3. Classification and Serialization tasks, in which the mind can cognitively distinguish various disambiguation criteria and group or order objects accordingly.

In terms of inference control this is the stage in which actual knowledge about how to control inference itself is first explored. This means an emerging understanding of inference itself as a cognitive task and methods for learning, which will be further developed in the following stages.

Also, in this stage a special cognitive task capability is gained: "Theory of Mind," which in cognitive science refers to the ability to understand the fact that not only oneself, but other sentient beings have memories, perceptions, and experiences. This is the ability to conceptually "put oneself in another's shoes" (even if you happen to assume incorrectly about them by doing so).

11.4.2.1 Conservation of Number

Conservation of number is an example of a learning problem classically categorized within Piaget's concrete-operational phase, a "conservation laws" problem, discussed in [Shu03] in the context of software that solves the problem using (logic-based and neural net) narrow-AI techniques. Conservation laws are very important to cognitive development.

Conservation is the idea that a quantity remains the same despite changes in appearance. If you show a child some objects and then spread them out, an infantile mind will focus on the spread, and believe that there are now more objects than before, whereas a concrete-operational mind will understand that the quantity of objects has not changed.

Conservation of number seems very simple, but from a developmental perspective it is actually rather difficult. "Solutions" like those given in [Shu03] that use neural networks or customized logical rule-bases to find specialized solutions that solve only this problem fail to fully address the issue, because these solutions don't create knowledge adequate to aid with the solution of related sorts of problems.

We hypothesize that this problem is hard enough that for an inference-based AGI system to solve it in a developmentally useful way, its inferences must be guided by meta-inferential lessons learned from prior similar problems. When approaching a number conservation problem, for example, a reasoning system might draw upon past experience with set-size problems (which may be trial-and-error experience). This is not a simple "machine learning" approach whose scope is restricted to the current problem, but rather a heuristically guided approach which (a) aggregates information from prior experience to guide solution formulation for the problem at hand, and (b) adds the present experience to the set of relevant information about quantification problems for future refinement of thinking.

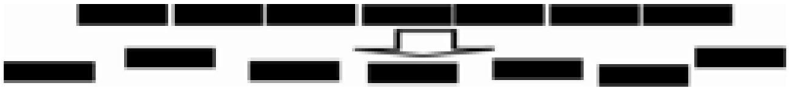


Fig. 11.6: Conservation of Number

For instance, a very simple context-specific heuristic that a system might learn would be: "When evaluating the truth value of a statement related to the number of objects in a set, it is generally not that useful to explore branches of the backwards-chaining search tree that contain relationships regarding the sizes, masses, or other physical properties of the objects in the set." This heuristic itself may go a long way toward guiding an inference process toward a correct solution to the problem–but it is not something that a mind needs to know "a priori." A concrete-operational stage mind may learn this by data-mining prior instances of inferences involving sizes of sets. Without such experience-based heuristics, the search tree for such a problem will likely be unacceptably large. Even if it is "solvable" without such heuristics, the solutions found may be overly fit to the particular problem and not usefully generalizable.

11.4.2.2 Theory of Mind

Consider this experiment: a preoperational child is shown her favorite "Dora the Explorer" DVD box. Asked what show she's about to see, she'll answer "Dora." However, when her parent plays the disc, it's "SpongeBob SquarePants." If you then ask her what show her friend will expect when given the "Dora" DVD box, she will respond "SpongeBob" although she just answered "Dora" for herself. A child lacking a theory of mind can not reason through what someone else would think given knowledge other than her own current knowledge. Knowledge of self is intrinsically related to the ability to differentiate oneself from others, and this ability may not be fully developed at birth.

Several theorists [BC94, Fod94], based in part on experimental work with autistic children, perceive theory of mind as embodied in an innate module of the mind activated at a certain developmental stage (or not, if damaged). While we consider this possible, we caution against adopting a simplistic view of the "innate vs. acquired" dichotomy: if there is innateness it may take the form of an innate predisposition to certain sorts of learning [EBJ+97].

Davidson [Dav84], Dennett [Den87] and others support the common belief that theory of mind is dependent upon linguistic ability. A major challenge to this prevailing philosophical stance came from Premack and Woodruff [PW78] who postulated that prelinguistic primates do indeed exhibit "theory of mind" behavior. While Premack and Woodruff's experiment itself has been challenged, their general result has been bolstered by follow-up work showing similar results such as [TC97]. It seems to us that while theory of mind depends on many of the same inferential capabilities as language learning, it is not intrinsically dependent on the latter.

There is a school of thought often called the *Theory Theory* [BW88, Car85, Wel90] holding that a child's understanding of mind is best understood in terms of the process of iteratively formulating and refuting a series of naive theories about others. Alternately, Gordon [Gor86] postulates that theory of mind is related to the ability to run cognitive simulations of others' minds using one's own mind as a model. We suggest that these two approaches are actually quite harmonious with one another. In an uncertain AGI context, both theories and simulations are grounded in collections of uncertain implications, which may be assembled in context-appropriate ways to form theoretical conclusions or to drive simulations. Even if there is a special "mind-simulator" dynamic in the human brain that carries out simulations of other minds in a manner fundamentally different from explicit inferential theorizing, the inputs to and the behavior of this simulator may take inferential form, so that the simulator is in essence a way of efficiently and implicitly producing uncertain inferential conclusions from uncertain premises.

We have thought through the details by CogPrime system should be able to develop theory of mind via embodied experience, though at time of writing practical learning experiments in this direction have not yet been done. We have not yet explored in detail the possibility of giving CogPrime a special, elaborately engineered "mind-simulator" component, though this would be possible; instead we have initially been pursuing a more purely inferential approach.

First, it is very simple for a CogPrime system to learn patterns such as "If I rotated by pi radians, I would see the yellow block." And it's not a big leap for PLN to go from this to the recognition that "You look like me, and you're rotated by pi radians relative to my orientation, therefore you probably see the yellow block." The only nontrivial aspect here is the "you look like me" premise.

Recognizing "embodied agent" as a category, however, is a problem fairly similar to recognizing "block" or "insect" or "daisy" as a category. Since the CogPrime agent can perceive most parts of its own "robot" body–its arms, its legs, etc.–it should be easy for the agent to figure out that physical objects like these look different depending upon its distance from them and its angle of observation. From this it should not be that difficult for the agent to understand that it is naturally grouped together with other embodied agents (like its teacher), not with blocks or bugs.

The only other major ingredient needed to enable theory of mind is "reflection"– the ability of the system to explicitly recognize the existence of knowledge in its own mind (note that this term "reflection" is not the same as our proposed "reflexive" stage of cognitive development). This exists automatically in CogPrime, via the built-in vocabulary of elementary procedures supplied for use within SchemaNodes (specifically, the atTime and TruthValue operators). Observing that "at time T, the weight of evidence of the link L increased from zero" is basically equivalent to observing that the link L was created at time T.

Then, the system may reason, for example, as follows (using a combination of several PLN rules including the above-given deduction rule):

Implication
    My eye is facing a block and it is not dark
    A relationship is created describing the block's color
Similarity
    My body
    My teacher's body
|-
Implication
    My teacher's eye is facing a block and it is not dark
    A relationship is created describing the block's color

This sort of inference is the essence of Piagetan "theory of mind." Note that in both of these implications the created relationship is represented as a variable rather than a specific relationship. The cognitive leap is that in the latter case the relationship actually exists in the teacher's implicitly hypothesized mind, rather than in CogPrime's mind. No explicit hypothesis or model of the teacher's mind need be created in order to form this implication–the hypothesis is created implicitly via inferential abstraction. Yet, a collection of implications of this nature may be used via an uncertain reasoning system like PLN to create theories and simulations suitable to guide complex inferences about other minds.

From the perspective of developmental stages, the key point here is that in a CogPrime context this sort of inference is too complex to be viably carried out via simple inference heuristics. This particular example must be done via forward chaining, since the big leap is to actually think of forming the implication that concludes inference. But there are simply too many combinations of relationships involving CogPrime's eye, body, and so forth for the PLN component to viably explore all of them via standard forward-chaining heuristics. Experience-guided heuristics are needed, such as the heuristic that if physical objects A and B are generally physically and functionally similar, and there is a relationship involving some part of A and some physical object R, it may be useful to look for similar relationships involving an analogous part of B and objects similar to R. This kind of heuristic may be learned by experience–and the masterful deployment of such heuristics to guide inference is what we hypothesize to characterize the concrete stage of development. The "concreteness" comes from the fact that inference control is guided by analogies to prior similar situations.

11.4.3 The Formal Stage

In the formal stage, as shown in Figure 11.7, an agent should be able to carry out arbitrarily complex inferences (constrained only by computational resources, rather than by fundamental restrictions on logical language or form) via including inference control as an explicit subject of abstract learning. Abstraction and inference about both the sensorimotor surround (world) and about abstract ideals themselves (including the final stages of indirect learning about inference itself) are fully developed.

Formal stage evaluation tasks are centered entirely around abstraction and higher-order inference tasks such as:

1. Mathematics and other formalizations.

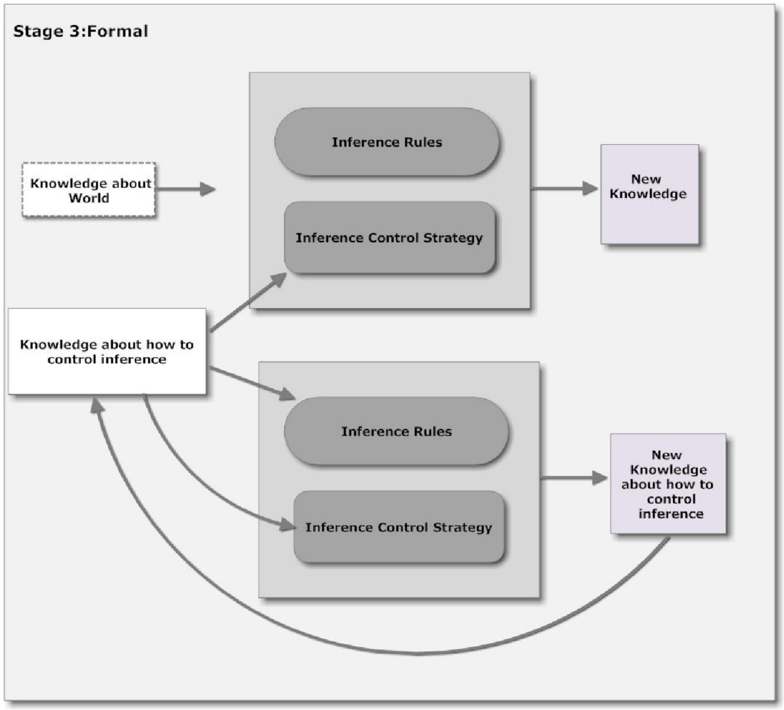

Stage 3:Formal
Knowledge about World
Inference Rules
Inference Control Strategy
New Knowledge
Knowledge about how to control inference
Inference Rules
Inference Control Strategy
New Knowledge about how to control inference

Fig. 11.7: Uncertain Inference in the Formal Stage

2. Scientific experimentation and other rigorous observational testing of abstract formalizations.
3. Social and philosophical modeling, and other advanced applications of empathy and the Theory of Mind.

In terms of inference control this stage sees not just perception of new knowledge about inference control itself, but inference controlled reasoning about that knowledge and the creation of abstract formalizations about inference control which are reasoned-upon, tested, and verified or debunked.

11.4.3.1 Systematic Experimentation

The Piagetan formal phase is a particularly subtle one from the perspective of uncertain inference. In a sense, AGI inference engines already have strong capability for formal reasoning built in. Ironically, however, no existing inference engine is capable of deploying its reasoning rules in a powerfully effective way, and this is because of the lack of inference control heuristics adequate for controlling abstract formal reasoning. These heuristics are what arise during Piaget's formal stage, and we propose that in the content of uncertain inference systems, they involve the application of inference itself to the problem of refining inference control.

A problem commonly used to illustrate the difference between the Piagetan concrete operational and formal stages is that of figuring out the rules for making pendulums swing quickly versus slowly [IP58]. If you ask a child in the formal stage to solve this problem, she may proceed to do a number of experiments, e.g. build a long string with a light weight, a long string with a heavy weight, a short string with a light weight and a short string with a heavy weight. Through these experiments she may determine that a short string leads to a fast swing, a long string leads to a slow swing, and the weight doesn't matter at all.

The role of experiments like this, which test "extreme cases," is to make cognition easier. The formal-stage mind tries to map a concrete situation onto a maximally simple and manipulable set of abstract propositions, and then reason based on these. Doing this, however, requires an automated and instinctive understanding of the reasoning process itself. The above-described experiments are good ones for solving the pendulum problem because they provide data that is very easy to reason about. From the perspective of uncertain inference systems, this is the key characteristic of the formal stage: formal cognition approaches problems in a way explicitly calculated to yield tractable inferences.

Note that this is quite different from saying that formal cognition involves abstractions and advanced logic. In an uncertain logic-based AGI system, even infantile cognition may involve these – the difference lies in the level of inference control, which in the infantile stage is simplistic and hard-wired, but in the formal stage is based on an understanding of what sorts of inputs lead to tractable inference in a given context.

11.4.4 The Reflexive Stage

In the reflexive stage (Figure 11.8), an intelligent agent is broadly capable of self-modifying its internal structures and dynamics.

As an example in the human domain: highly intelligent and self-aware adult humans may carry out reflexive cognition by explicitly reflecting upon their own inference processes and trying to improve them. An example is the intelligent improvement of uncertain-truth-value-manipulation formulas. It is well demonstrated that even educated humans typically make numerous errors in probabilistic reasoning [GGK02]. Most people don't realize it and continue to systematically make these errors throughout their lives. However, a small percentage of individuals make an explicit effort to increase their accuracy in making probabilistic judgments by consciously endeavoring to internalize the rules of probabilistic inference into their automated cognition processes.

In the uncertain inference based AGI context, what this means is: In the reflexive stage an entity is able to include inference control itself as an explicit subject of abstract learning (i.e. the ability to reason about one's own tactical and strategic approach to modifying one's own learning and thinking), and modify these inference control strategies based on analysis of experience with various cognitive approaches.

Ultimately, the entity can self-modify its internal cognitive structures. Any knowledge or heuristics can be revised, including metatheoretical and metasystemic thought itself. Initially this is done indirectly, but at least in the case of AGI systems it is theoretically possible to also do so directly. This might be considered as a separate stage of Full Self Modification, or else as the end phase of the reflexive stage. In the context of logical reasoning, self modification of inference control itself is the primary task in this stage. In terms of inference control this

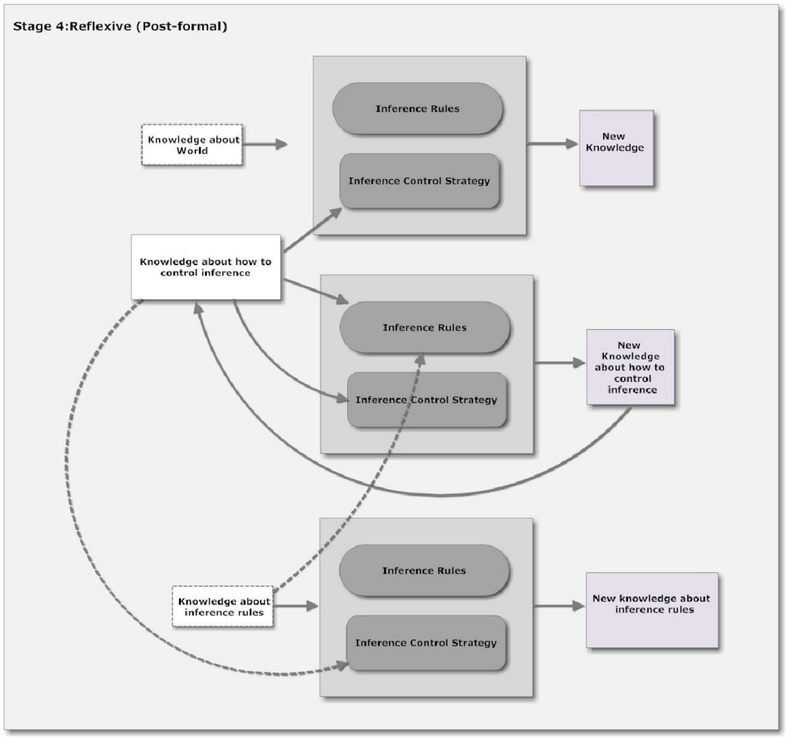

Stage 4:Reflexive (Post-formal)
Knowledge about World
Inference Rules
Inference Control Strategy
New Knowledge
Knowledge about how to control inference
Inference Rules
Inference Control Strategy
New Knowledge about how to control inference
Knowledge about inference rules
Inference Rules
Inference Control Strategy
New knowledge about inference rules

Fig. 11.8: The Reflexive Stage

stage adds an entire new feedback loop for reasoning about inference control itself, as shown in Figure 11.8.

As a very concrete example, in later chapters we will see that, while PLN is founded on probability theory, it also contains a variety of heuristic assumptions that inevitably introduce a certain amount of error into its inferences. For example, PLN's probabilistic deduction embodies a heuristic independence assumption. Thus PLN contains an alternate deduction formula called the "concept geometry formula" that is better in some contexts, based on the assumption that ConceptNodes embody concepts that are roughly spherically-shaped in attribute space. A highly advanced CogPrime system could potentially augment the independence-based and concept-geometry-based deduction formulas with additional formulas of its own derivation, optimized to minimize error in various contexts. This is a simple and straightforward example of reflexive cognition – it illustrates the power accessible to a cognitive system that has formalized and reflected upon its own inference processes, and that possesses at least some capability to modify these.

In general, AGI systems can be expected to have much broader and deeper capabilities for self-modification than human beings. Ultimately it may make sense to view the AGI systems we implement as merely "initial conditions" for ongoing self-modification and self-organization. Chapter **??** discusses some of the potential technical details underlying this sort of thoroughgoing AGI self-modification.

Chapter 12
The Engineering and Development of Ethics

Co-authored with Stephan Vladimir Bugaj and Joel Pitt

12.1 Introduction

Most commonly, if a work on advanced AI mentions ethics at all, it occurs in a final summary chapter, discussing in broad terms some of the possible implications of the technical ideas presented beforehand. It's no coincidence that the order is reversed here: in the case of CogPrime, AGI-ethics considerations played a major role in the design process ... and thus the chapter on ethics occurs near the beginning rather than the end. In the CogPrime approach, ethics is not a particularly distinct topic, being richly interwoven with cognition and education and other aspects of the AGI project.

The ethics of advanced AGI is a complex issue with multiple aspects. Among the many issues there are:

1. Risks posed by the possibility of human beings using AGI systems for evil ends
2. Risks posed by AGI systems created without well-defined ethical systems
3. Risks posed by AGI systems with initially well-defined and sensible ethical systems eventually going rogue – an especially big risk if these systems are more generally intelligent than humans, and possess the capability to modify their own source code
4. the ethics of experimenting on AGI systems when one doesn't understand the nature of their experience
5. AGI rights: in what circumstances does using an AGI as a tool or servant constitute "slavery"

In this chapter we will focus mainly (though not exclusively) on the question of *how to create an AGI with a rational and beneficial ethical system*. After a somewhat wide-ranging discussion, we will conclude with eight general points that we believe should be followed in working toward "Friendly AGI" – most of which have to do, not with the internal design of the AGI, but with the way the AGI is taught and interfaced with the real world.

While most of the particulars discussed in this book have nothing to do with ethics, it's important for the reader to understand that AGI-ethics considerations have played a major role in many of our design decisions, underlying much of the technical contents of the book. As the materials in this chapter should make clear, ethicalness is probably not something that one can meaningfully tack onto an AGI system at the end, after developing the rest – it is likely infeasible to architect an intelligent agent and then add on an "ethics module." Rather, ethics is something that has to do with all the different memory systems and cognitive processes that constitute an intelligent system – and it's something that involves both cognitive architecture *and* the exploration a system does and the instruction it receives. It's a very complex matter that is richly intermixed with all the other aspects of intelligence, and here we will treat it as such.

12.2 Review of Current Thinking on the Risks of AGI

Before proceeding to outline our own perspective on AGI ethics in the context of CogPrime, we will review the main existing strains of thought on the potential ethical dangers associated with AGI. One science fiction film after another has highlighted these dangers, lodging the issue deep in our cultural awareness; unsurprisingly, much less attention has been paid to serious analysis of the risks in their various dimensions, but there is still a non-trivial literature worth paying attention to.

Hypothetically, an AGI with superhuman intelligence and capability could dispense with humanity altogether – i.e. posing an "existential risk" [Bos02]. In the worst case, an evil but brilliant AGI, perhaps programmed by a human sadist, could consign humanity to unimaginable tortures (i.e. realizing a modern version of the medieval Christian visions of hell). On the other hand, the potential benefits of powerful AGI also go literally beyond human imagination. It seems quite plausible that an AGI with massively superhuman intelligence and positive disposition toward humanity could provide us with truly dramatic benefits, such as a virtual end to material scarcity, disease and aging. Advanced AGI could also help individual humans grow in a variety of directions, including directions leading beyond "legacy humanity," according to their own taste and choice.

Eliezer Yudkowsky has introduced the term "Friendly AI", to refer to advanced AGI systems that act with human benefit in mind [Yud06]. Exactly what this means has not been specified precisely, though informal interpretations abound. Goertzel [Goe06b] has sought to clarify the notion in terms of three core values of Joy, Growth and Freedom. In this view, a Friendly AI would be one that advocates individual and collective human joy and growth, while respecting the autonomy of human choices.

Some (for example, Hugo de Garis, [DG05]), have argued that Friendly AI is essentially an impossibility, in the sense that the odds of a dramatically superhumanly intelligent mind worrying about human benefit are vanishingly small. If this is the case, then the best options for the human race would presumably be to either avoid advanced AGI development altogether, or to else fuse with AGI before it gets too strongly superhuman, so that beings-originated-as-humans can enjoy the benefits of greater intelligence and capability (albeit at cost of sacrificing their humanity).

Others (e.g. Mark Waser [Was09]) have argued that Friendly AI is essentially inevitable, because greater intelligence correlates with greater morality. Evidence from evolutionary and human history is adduced in favor of this point, along with more abstract arguments.

Yudkowsky [Yud06] has discussed the possibility of creating AGI architectures that are in some sense "provably Friendly" – either mathematically, or else at least via very tight lines of rational verbal argumentation. However, several issues have been raised with this approach. First, it seems likely that proving mathematical results of this nature would first require dramatic advances in multiple branches of mathematics. Second, such a proof would require a formalization of the goal of "Friendliness," which is a subtler matter than it might seem [Leg06b, Leg06a].

Formalization of human morality has vexed moral philosophers for quite some time. Finally, it is unclear the extent to which such a proof could be created in a generic, environment-independent way – but if the proof depends on properties of the physical environment, then it would require a formalization of the environment itself, which runs up against various problems such as the complexity of the physical world and also the fact that we currently have no complete, consistent theory of physics. Kaj Sotala has provided a list of 14 objections to the Friendly AI concept, and suggested answers to each of them [Sot11]. Stephen Omohundro [Omo08] has argued that any advanced AI system will very likely demonstrate certain "basic AI drives", such as desiring to be rational, to self-protect, to acquire resources, and to preserve and protect its utility function and avoid counterfeit utility; these drives, he suggests, must be taken carefully into account in formulating approaches to Friendly AI.

The problem of formally or at least very carefully defining the goal of Friendliness has been considered from a variety of perspectives, none showing dramatic success. Yudkowsky [Yud04] has suggested the concept of "Coherent Extrapolated Volition", which roughly refers to the extrapolation of the common values of the human race. Many subtleties arise in specifying this concept – e.g. if Bob Jones is often possessed by a strong desire to kill all Martians, but he deeply aspires to be a nonviolent person, then the CEV approach would not rate "killing Martians" as part of Bob's contribution to the CEV of humanity.

Goertzel [Goe10a] has proposed a related notion of Coherent Aggregated Volition (CAV), which eschews the subtleties of extrapolation, and simply seeks a reasonably *compact, coherent, consistent* set of values that is fairly close to the collective value-set of humanity. In the CAV approach, "killing Martians" would be removed from humanity's collective value-set because it's uncommon and not part of the most compact/coherent/consistent overall model of human values, rather than because of Bob Jones' aspiration to nonviolence.

One thought we have recently entertained is that the core concept underlying CAV might be better thought of as CBV or "Coherent Blended Volition." CAV seems to be easily misinterpreted as meaning the average of different views, which was not the original intention. The CBV terminology clarifies that the CBV of a diverse group of people should not be thought of as an average of their perspectives, but as something more analogous to a "conceptual blend" [FT02] – incorporating the most essential elements of their divergent views into a whole that is overall compact, elegant and harmonious. The subtlety here (to which we shall return below) is that for a CBV blend to be broadly acceptable, the different parties whose views are being blended must agree to some extent that enough of the essential elements of their own views have been included. The process of arriving at this sort of consensus may involve extrapolation of a roughly similar sort to that considered in CEV.

Multiple attempts at axiomatization of human values have also been attempted, e.g. with a view toward providing near-term guidance to military robots (see e.g. Arkin's excellent though chillingly-titled book *Governing Lethal Behavior in Autonomous Robots* [Ark09b], the result of US military funded research). However, there are reasonably strong arguments that human values (similarly to e.g. human language or human perceptual classification rules) are too complex and multifaceted to be captured in any compact set of formal logic rules. Wallach [WA10] has made this point eloquently, and argued the necessity of fusing top-down (e.g. formal logic based) and bottom-up (e.g. self-organizing learning based) approaches to machine ethics.

A number of more sociological considerations also arise. It is sometimes argued that the risk from highly-advanced AGI going morally awry on its own may be less than that of moderately-advanced AGI being used by human beings to advocate immoral ends. This possibility gives rise to questions about the ethical value of various practical modalities of AGI development, for instance:

- Should AGI be developed in a top-secret installation by a select group of individuals selected for a combination of technical and scientific brilliance and moral uprightness, or other qualities deemed relevant (a "closed approach")? Or should it be developed out in the open, in the manner of open-source software projects like Linux? (an "open approach"). The open approach allows the collective intelligence of the world to more fully participate – but also potentially allows the more unsavory elements of the human race to take some of the publicly-developed AGI concepts and tools private, and develop them into AGIs with selfish or evil purposes in mind. Is there some meaningful intermediary between these extremes?
- Should governments regulate AGI, with Friendliness in mind (as advocated carefully by e.g Bill Hibbard [Hib02])? Or will this just cause AGI development to move to the handful of countries with more liberal policies? ... or cause it to move underground, where nobody can see the dangers developing? As a rough analogue, it's worth noting that the US government's imposition of restrictions on stem cell research, under President George W. Bush, appears to have directly stimulated the provision of additional funding for stem cell research in other nations like Korea, Singapore and China.

The former issue is, obviously, highly relevant to CogPrime (which is currently being developed via the open source CogPrime project); and so the various dimensions of this issues are worth briefly sketching here.

We have a strong skepticism of self-appointed elite groups that claim (even if they genuinely believe) that they know what's best for everyone, and a healthy respect for the power of collective intelligence and the Global Brain, which the open approach is ideal for tapping. On the other hand, we also understand the risk of terrorist groups or other malevolent agents forking an open source AGI project and creating something terribly dangerous and destructive. Balancing these factors against each other rigorously, seems beyond the scope of current human science.

Nobody really understands the social dynamics by which open technological knowledge plays out in our *current* world, let alone hypothetical future scenarios. Right now there exists open knowledge about many very dangerous technologies, and there exist many terrorist groups, yet these groups fortunately make scant use of these technologies. The reasons why appear to be essentially sociological – the people involved in these terrorist groups tend not to be the ones who have mastered the skills of turning public knowledge on cutting-edge technologies into real engineered systems. But while it's easy to observe this sociological phenomenon, we certainly have no way to estimate its *quantitative extent* from first principles. We don't really have a strong understanding of how safe we are right now, given the technology knowledge available right now via the Internet, textbooks, and so forth. Even relatively straightforward issues such as nuclear proliferation remain confusing, even to the experts.

It's also quite clear that keeping powerful AGI locked up by an elite group doesn't really provide reliable protection against malevolent human agents. History is rife with such situations going awry, e.g. by the leadership of the group being subverted, or via brute force inflicted by some outside party, or via a member of the elite group defecting to some outside group in the interest of personal power or reward or due to group-internal disagreements, etc. There are many things that can go wrong in such situations, and the confidence of any particular group that they are immune to such issues, cannot be taken very seriously. Clearly, neither the open nor closed approach qualifies as a panacea.

12.3 The Value of an Explicit Goal System

One of the subtle issues confronted in the quest to design ethical AGIs is how closely one wants to emulate human ethical judgment and behavior. Here one confronts the brute fact that, even according to their own deeply-held standards, humans are not all that ethical. One high-level conclusion we came to very early in the process of designing CogPrime is that, just as humans are not the most intelligent minds achievable, they are also not the most ethical minds achievable. Even if one takes human ethics, broadly conceived, as the standard – there are almost surely possible AGI systems that are much more ethical *according to human standards* than nearly all human beings. This is not mainly because of ethics-specific features of the human mind, but rather because of the nature of the human motivational system, which leads to many complexities that drive humans to behaviors that are unethical according to their own standards. So, one of the design decisions we made for CogPrime – with ethics as well as other reasons in mind – was *not* to closely imitate the human motivational system, but rather to craft a novel motivational system combining certain aspects of the human motivational system with other profoundly non-human aspects.

On the other hand, the design of ethical AGI systems still has a lot to gain from the study of human ethical cognition and behavior. Human ethics has many aspects, which we associate here with the different types of memory, and it's important that AGI systems can encompass all of them. Also, as we will note below, human ethics develops in childhood through a series of natural stages, parallel to and entwined with the cognitive developmental stages reviewed in Chapter 11 above. We will argue that for an AGI with a virtual or robotic body, it makes sense to think of ethical development as proceeding through similar stages. In a CogPrime context, the particulars of these stages can then be understood in terms of the particulars of CogPrime's cognitive processes – which brings AGI ethics from the domain of theoretical abstraction into the realm of practical algorithm design and education.

But even if the human stages of ethical development make sense for non-human AGIs, this doesn't mean the particulars of the human motivational system need to be replicated in these AGIs, regarding ethics or other matters. A key point here is that, in the context of human intelligence, the concept of a "goal" is a descriptive abstraction. But in the AGI context, it seems quite valuable to introduce goals as explicit design elements (which is what is done in CogPrime) – both for ethical reasons and for broader AGI design reasons.

Humans may adopt goals for a time and then drop them, may pursue multiple conflicting goals simultaneously, and may often proceed in an apparently goal-less manner. Sometimes the goal that a person appears to be pursuing, may be very different than the one they think they're pursuing. Evolutionary psychology [BDL93] argues that, directly or indirectly, all humans are ultimately pursuing the goal of maximizing the inclusive fitness of their genes – but given the complex mix of evolution and self-organization in natural history [Sal93], this is hardly a general explanation for human behavior. Ultimately, in the human context, "goal" is best thought of as a frequently useful heuristic concept.

AGI systems, however, need not emulate human cognition in every aspect, and may be architected with explicit "goal systems." This provides no guarantee that said AGI systems will actually pursue the goals that their goal systems specify – depending on the role that the goal system plays in the overall system dynamics, sometimes other dynamical phenomena might intervene and cause the system to behave in ways opposed to its explicit goals. However, we submit that this design sketch provides a better framework than would exist in an AGI system closely emulating the human brain.

We realize this point may be somewhat contentious – a counter-argument would be that the human brain is known to support at least *moderately* ethical behavior, according to human ethical standards, whereas less brain-like AGI systems are much less well understood. However, the obvious counter-counterpoints are that:

- Humans are not all that consistently ethical, so that creating AGI systems potentially much more practically powerful than humans, but with closely humanlike ethical, motivational and goal systems, could in fact be quite dangerous
- The effect on a human-like ethical/motivational/goal system of increasing the intelligence, or changing the physical embodiment or cognitive capabilities, of the agent containing the system, is unknown and difficult to predict given all the complexities involved

The course we tentatively recommend, and are following in our own work, is to develop AGI systems with explicit, hierarchically-dominated goal systems. That is:

- create one or more "top goals" (we call them Ubergoals in CogPrime)
- have the system derive subgoals from these, using its own intelligence, potentially guided by educational interaction or explicit programming
- have a significant percentage of the system's activity governed by the explicit pursuit of these goals

Note that the "significant percentage" need not be 100%; CogPrime, for example, combines explicitly goal-directed activity with other "spontaneous" activity. Requiring that all activity be explicitly goal-directed may be too strict a requirement to place on AGI architectures.

The next step, of course, is for the top-level goals to be chosen in accordance with the principle of human-Friendliness. The next one of our eight points, about the Global Brain, addresses one way of doing this. In our near-term work with CogPrime, we are using simplistic approaches, with a view toward early-stage system testing.

12.4 Ethical Synergy

An explicit goal system provides an explicit way to ensure that ethical principles (as represented in system goals) play a significant role in guiding an AGI system's behavior. However, in an integrative design like CogPrime the goal system is only a small part of the overall story, and it's important to also understand how ethics relates to the other aspects of the cognitive architecture.

One of the more novel ideas presented in this chapter is that different types of ethical intuition may be associated with different types of memory – and to possess mature ethics, a mind must display *ethical synergy* between the ethical processes associated with its memory types. Specifically, we suggest that:

- **Episodic memory** corresponds to the process of ethically assessing a situation based on similar prior situations
- **Sensorimotor memory** corresponds to "mirror neuron" type ethics, where you feel another person's feelings via mirroring their physiological emotional responses and actions
- **Declarative memory** corresponds to rational ethical judgment

- **Procedural memory** corresponds to "ethical habit" ... learning by imitation and reinforcement to do what is right, even when the reasons aren't well articulated or understood
- **Attentional memory** corresponds to the existence of appropriate patterns guiding one to pay adequate attention to ethical considerations at appropriate times
- **Intentional memory** corresponds to the pervasion of ethics through one's choices about subgoaling (which leads into "when do the ends justify the means" ethical-balance questions)

One of our suggestions regarding AGI ethics is that an ethically mature person or AGI must both master and balance all these kinds of ethics. We will focus especially here on declarative ethics, which corresponds to Kohlberg's theory of logical ethical judgment; and episodic ethics, which corresponds to Gilligan's theory of empathic ethical judgment. Ultimately though, all five aspects are critically important; and a CogPrime system if appropriately situated and educated should be able to master and integrate all of them.

12.4.1 Stages of Development of Declarative Ethics

Complementing generic theories of cognitive development such as Piaget's and Perry's, theorists have also proposed specific stages of moral and ethical development. The two most relevant theories in this domain are those of Kohlberg and Gilligan, which we will review here, both individually and in terms of their integration and application in the AGI context.

Lawrence Kohlberg's [KLH83, Koh81] moral development model, called the "ethics of justice" by Gilligan, is based on a rational modality as the central vehicle for moral development. In our perspective this is a firmly *declarative* form of ethics, based on explicit analysis and reasoning. It is based on an impartial regard for persons, proposing that ethical consideration must be given to all individual intelligences without a priori judgment (prejudice). Consideration is given for individual merit and preferences, and the goals of an ethical decision are equal treatment (in the general, not necessarily the particular) and reciprocity. Echoing Kant's [Kan64] categorical imperative, the decisions considered most successful in this model are those which exhibit "reversibility", where a moral act within a particular situation is evaluated in terms of whether or not the act would be satisfactory even if particular persons were to switch roles within the situation. In other words, a situational, contextualized "do unto others as you would have them do unto you" criterion. The ethics of justice can be viewed as three stages (each of which has six substages, on which we will not elaborate here), depicted in Table 12.1.

In Kohlberg's perspective, cognitive development level contributes to moral development, as moral understanding emerges from increased cognitive capability in the area of ethical decision making in a social context. Relatedly, Kohlberg also looks at stages of social perspective and their consequent interpersonal outlook. As shown in Table 12.1, these are correlated to the stages of moral development, but also map onto Piagetian models of cognitive development (as pointed out e.g. by Gibbs [Gib78], who presents a modification/interpretation of Kohlberg's ideas intended to align them more closely with Piaget's). Interpersonal outlook can be understood as rational understanding of the psychology of other persons (a theory of mind, with or without empathy). Stage One, emergent from the infantile congitive stage, is entirely selfish as only self awareness has developed. As cognitive sophistication about ethical considerations increases, so do the moral and social perspective stages. Concrete and formal cognition bring about the first instrumental egoism, and then social relations and systems perspectives, and

Stage	**Substages**
Pre-Conventional	• Obedience and Punishment Orientation • Self-interest orientation
Conventional	• Interpersonal accord (conformity) orientation • Authority and social-order maintaining (law and order) orientation
Post-Conventional	• Social contract (human rights) orientation • Universal ethical principles (universal human rights) orientation

Table 12.1: Kohlberg's Stages of Development of the Ethics of Justice

from formal and then reflexive thinking about ethics comes the post-conventional modalities of contractualism and universal mutual respect.

Stage of Social Perspective	**Interpersonal Outlook**
Blind egoism	No interpersonal perspective. Only self is considered.
Instrumental egoism	See that others have goals and perspectives, and either conform to or rebel against norms.
Social Relationships perspective	Able to see abstract normative systems
Social Systems perspective	Recognize positive and negative intentions
Contractual perspective	Recognize that contracts (mutually beneficial agreements of any kind) will allow intelligences to increase the welfare of both.
Universal principle of mutual respect	See how human fallibility and frailty are impacted by communication.

Table 12.2: Kohlberg's Stages of Development of Social Perspective and Interpersonal Morals

12.4.1.1 Uncertain Inference and the Ethics of Justice

Taking our cue from the analysis given in Chapter 11 of Piagetan stages in uncertain inference based AGI systems (such as CogPrime), we may explore the manifestation of Kohlberg's stages in AGI systems of this nature. Uncertain inference seems generally well-suited as a declarative-ethics learning system, due to the nuanced ethical environment of real world situations. Probabilistic knowledge networks can model belief networks, imitative reinforcement learning based ethical pedagogy, and even simplistic moral maxims. In principle, they have the flexibility to deal with complex ethical decisions, including not only weighted "for the greater good" dichotomous decision making, but also the ability to develop moral decision networks which do not require that all situations be solved through resolution of a dichotomy.

When more than one person is being affected by an ethical decision, making a decision based on reducing two choices to a single decision can often lead to decisions of dubious ethics. However, a sufficiently complex uncertain inference network can represent alternate choices in which multiple actions are taken that have equal (or near equal) belief weight but have very different particulars – but because the decisions are applied in different contexts (to different groups of individuals) they are morally equivalent. Though each individual action appears equally believable, were any single decision applied to the entire population one or more individual may be harmed, and the morally superior choice is to make case-dependent decisions. Equal moral treatment is a general principle, and too often the mistake is made by thinking that to achieve this general principle the particulars must be equal. This is not the case. Different treatment of different individuals can result in morally equivalent treatment of all involved individuals, and may be vastly morally superior to treating all the individuals with equal particulars. Simply taking the largest population and deciding one course of action based on the result that is most appealing to that largest group is not generally the most moral action.

Uncertain inference, especially a complex network with high levels of resource access as may be found in a sophisticated AGI, is well suited for complex decision making resulting in a multitude of actions, and of analyzing the options to find the set of actions that are ethically optimal particulars for each decision context. Reflexive cognition and post-commitment moral understanding may be the goal stages of an AGI system, or any intelligence, but the other stages will be passed through on the way to that goal, and realistically some minds will never reach higher order cognition or morality with regards to any context, and others will not be able to function at this high order in every context (all currently known minds fail to function at the highest order cognitively or morally in some contexts).

Infantile and concrete cognition are the underpinnings of the egoist and socialized stages, with formal aspects also playing a role in a more complete understanding of social models when thinking using the social modalities. Cognitively infantile patterns can produce no more than blind egoism as without a theory of mind, there is no capability to consider the other. Since most intelligences acquire concrete modality and therefore some nascent social perspective relatively quickly, most egoists are instrumental egoists. The social relationship and systems perspectives include formal aspects which are achieved by systematic social experimentation, and therefore experiential reinforcement learning of correct and incorrect social modalities. Initially this is a one-on-one approach (relationship stage), but as more knowledge of social action and consequences is acquired, a formal thinker can understand not just consequentiality but also intentionality in social action.

Extrapolation from models of individual interaction to general social theoretic notions is also a formal action. Rational, logical positivist approaches to social and political ideas, however, are the norm of formal thinking. Contractual and committed moral ethics emerges from a higher-order formalization of the social relationships and systems patterns of thinking. Generalizations of social observation become, through formal analysis, systems of social and political doctrine. Highly committed, but grounded and logically supportable, belief is the hallmark of formal cognition as expressed contractual moral stage. Though formalism is at work in the socialized moral stages, its fullest expression is in committed contractualism.

Finally, reflexive cognition is especially important in truly reaching the post-commitment moral stage in which nuance and complexity are accommodated. Because reflexive cognition is necessary to change one's mind not just about particular rational ideas, but whole *ways of* *thinking*, this is a cognitive precedent to being able to reconsider an entire belief system, one that has had contractual logic built atop reflexive adherence that began in early development. If the initial moral system is viewed as positive and stable, then this cognitive capacity is seen as dangerous and scary, but if early morality is stunted or warped, then this ability is seen as enlightened. However, achieving this cognitive stage does not mean one automatically changes their belief systems, but rather that the mental machinery is in place to consider the possibilities. Because many people do not reach this level of cognitive development in the area of moral and ethical thinking, it is associated with negative traits ("moral relativism" and "flip-flopping"). However, this cognitive flexibility generally leads to more sophisticated and applicable moral codes, which in turn leads to morality which is actually more stable because it is built upon extensive and deep consideration rather than simple adherence to reflexive or rationalized ideologies.

12.4.2 Stages of Development of Empathic Ethics

Complementing Kohlberg's logic-and-justice-focused approach, Carol Gilligan's [Gil82] "ethics of care" model is a moral development theory which posits that empathetic understanding plays the central role in moral progression from an initial self-centered modality to a socially responsible one. The ethics of care model is concerned with the ways in which an individual cares (responds to dilemmas using empathetic responses) about self and others. As shown in Table 12.3, the ethics of care is broken into the same three primary stage as Kohlberg, but with a focus on empathetic, emotional caring rather than rationalized, logical principles of justice.

Stage	**Principle of Care**
Pre-Conventional	Individual Survival
Conventional	Self Sacrifice for the Greater Good
Post-Conventional	Principle of Nonviolence (do not hurt others, or oneself)

Table 12.3: Gilligan's Stages of the Ethics of Care

For an "ethics of care" approach to be applied in an AGI, the AGI must be capable of internal simulation of other minds it encounters, in a similar manner to how humans regularly simulate one another internally. Without any mechanism for internal simulation, it is unlikely that an AGI can develop any sort of empathy toward other minds, as opposed to merely logically or probabilistically modeling other agents' behavior or other minds' internal contents. In a CogPrime context, this ties in closely with how CogPrime handles episodic knowledge – partly via use of an internal simulation world, which is able to play "mental movies" of prior and hypothesized scenarios within the AGI system's mind.

However, in humans empathy involves more than just simulation, it also involves sensorimotor responses, and of course emotional responses – a topic we will discuss in more depth in Appendix **??** where we review the functionality of mirror neurons and mirror systems in the human brains. When we see or hear someone suffering, this sensory input causes motor responses in us similar to if we were suffering ourselves, which initiates emotional empathy and corresponding cognitive processes.

Thus, empathic "ethics of care" involves a combination of episodic and sensorimotor ethics, complementing the mainly declarative ethics associated with the "ethics of justice."

In Gilligan's perspective, the earliest stage of ethical development occurs before empathy becomes a consistent and powerful force. Next, the hallmark of the conventional stage is that at this point, the individual is so overwhelmed with their empathic response to others that they neglect themselves in order to avoid hurting others. Note that this stage doesn't occur in Kohlberg's hierarchy at all. Kohlberg and Gilligan both begin with selfish unethicality, but their following stages diverge. A person could in principle manifest Gilligan's conventional stage without having a refined sense of justice (thus not entering Kohlberg's conventional stage); or they could manifest Kohlberg's conventional stage without partaking in an excessive degree of self-sacrifice (thus not entering Gilligan's conventional stage). We will suggest below that in fact the empathic and logical aspects of ethics are more unified in real human development than these separate theories would suggest. However, even if this is so, the possibility is still there that in some AGI systems the levels of declarative and empathic ethics could wildly diverge.

It is interesting to note that Gilligan's and Kohlberg's final stages converge more closely than their intermediate ones. Kohlberg's post-conventional stage focuses on universal rights, and Gilligan's on universal compassion. Still, the foci here are quite different; and, as will be elaborated below, we believe that both Kohlberg's and Gilligan's theories constitute very partial views of the actual end-state of ethical advancement.

12.4.3 An Integrative Approach to Ethical Development

We feel that both Kohlberg's and Gilligan's theories contain elements of the whole picture of ethical development, and that both approaches are necessary to create a moral, ethical artificial general intelligence – just as, we suggest, both internal simulation and uncertain inference are necessary to create a sufficiently intelligent and volitional intelligence in the first place. Also, we contend, the lack of direct analysis of the underlying psychology of the stages is a deficiency shared by both the Kohlberg and Gilligan models as they are generally discussed. A successful model of integrative ethics necessarily contains elements of both the care and justice models, as well as reference to the underlying developmental psychology and its influence on the character of the ethical stage. Furthermore, intentional and attentional ethics need to be brought into the picture, complementing Kohlberg's focus on declarative knowledge and Gilligan's focus on episodic and sensorimotor knowledge.

With these notions in mind, we propose the following integrative theory of the stages of ethical development, shown in Tables 12.4, 12.5 and 12.6. In our integrative model, the justice-based and empathic aspects of ethical judgment are proposed to develop together. Of course, in any one individual, one or another aspect may be dominant. Even so, however, the combination of the two is equally important as either of the two individual ingredients.

For instance, we suggest that in any psychologically healthy human, the conventional stage of ethics (typifying childhood, and in many cases adulthood as well) involves a combination of Gilligan-esqe empathic ethics and Kohlberg-esque ethical reasoning. This combination is supported by Piagetan concrete operational cognition, which allows moderately sophisticated linguistic interaction, theory of mind, and symbolic modeling of the world.

And, similarly, we propose that in any truly ethically mature human, empathy and rational justice are both fully developed. Indeed the two interpenetrate each other deeply.

Once one goes beyond simplistic, childlike notions of fairness ("an eye for an eye" and so forth), applying rational justice in a purely intellectual sense is just as difficult as any other real-world logical inference problem. Ethical quandaries and quagmires are easily encountered, and are frequently cut through by a judicious application of empathic simulation.

On the other hand, empathy is a far more powerful force when used in conjunction with reason: analogical reasoning lets us empathize with situations we have never experienced. For instance, a person who has never been clinically depressed may have a hard time empathizing with individuals who are; but using the power of reason, they can imagine their worst state of depression magnified by several times and then extended over a long period of time, and then reason about what this might be like ... and empathize based on their inferential conclusion. Reason is not antithetical to empathy but rather is the key to making empathy more broadly impactful.

Finally, the enlightened stage of ethical development involves both a deeper compassion and a more deeply penetrating rationality and objectiveness. Empathy with all sentient beings is manageable in everyday life only once one has deeply reflected on one's own self and largely freed oneself of the confusions and illusions that characterize much of the ordinary human's inner existence. It is noteworthy, for example, that Buddhism contains both a richly developed ethics of universal compassion, and also an intricate logical theory of the inner workings of cognition [Stc00], detailing in exquisite rational detail the manner in which minds originate structures and dynamics allowing them to comprehend themselves and the world.

12.4.4 Integrative Ethics and Integrative AGI

What does our integrative approach to ethical development have to say about the ethical development of AGI systems? The lessons are relatively straighforward, if one considers an AGI system that, like CogPrime, explicitly contains components dedicated to logical inference and to simulation. Application of the above ethical ideas to other sorts of AGI systems is also quite possible, but would require a lengthier treatment and so won't be addressed here.

In the context of a CogPrime-type AGI system, Kolhberg's stages correspond to increasingly sophisticated application of logical inference to matters of rights and fairness. It is not clear whether humans contain an innate sense of fairness. In the context of AGIs, it would be possible to explicitly wire a sense of fairness into an AGI system, but in the context of a rich environment and active human teachers, this actually appears quite unnecessary. Experiential instruction in the notions of rights and fairness should suffice to teach an inference-based AGI system how to manipulate these concepts, analogously to teaching the same AGI system how to manipulate number, mass and other such quantities. Ascending the Kohlberg stages is then mainly a matter of acquiring the ability to carry out suitably complex inferences in the domain of rights and fairness. The hard part here is inference control – choosing which inference steps to take – and in a sophisticated AGI inference engine, inference control will be guided by experience, so that the more ethical judgments the system has executed and witnessed, the better it will become at making new ones. And, as argued above, simulative activity can be extremely valuable for aiding with inference control. When a logical inference process reaches a point of acute uncertainty (the backward or forward chaining inference tree can't decide which expansion step to take), it can run a simulation to cut through the confusion – i.e., it can use empathy to decide which

Stage	**Characteristics**
Pre-ethical	• Piagetan infantile to early concrete (aka pre-operational) • Radical selfishness or selflessness may, but do not necessarily, occur • No coherent, consistent pattern of consideration for the rights, intentions or feelings of others • Empathy is generally present, but erratically
Conventional Ethics	• Concrete cognitive basis • Perry's Dualist and Multiple stages • The common sense of the Golden Rule is appreciated, with cultural conventions for abstracting principles from behaviors • One's own ethical behavior is explicitly compared to that of others • Development of a functional, though limited, theory of mind • Ability to intuitively conceive of notions of fairness and rights • Appreciation of the concept of law and order, which may sometimes manifest itself as systematic obedience or systematic disobedience • Empathy is more consistently present, especially with others who are directly similar to oneself or in situations similar to those one has directly experienced • Degrees of selflessness or selfishness develop based on ethical groundings and social interactions.

Table 12.4: Integrative Model of the Stages of Ethical Development, Part 1

logical inference step to take in thinking about applying the notions of rights and fairness to a given situation.

Gilligan's stages correspond to increasingly sophisticated control of empathic simulation – which in a CogPrime-type AGI system, is carried out by a specific system component devoted to running internal simulations of aspects of the outside world, which includes a subcomponent specifically tuned for simulating sentient actors. The conventional stage has to do with the raw, uncontrolled capability for such simulation; and the post-conventional stage corresponds to its contextual, goal-oriented control. But controlling empathy, clearly, requires subtle management of various uncertain contextual factors, which is exactly what uncertain logical inference is good at – so, in an AGI system combining an uncertain inference component with a simulative component, it is the inference component that would enable the nuanced control of empathy allowing the ascent to Gilligan's post-conventional stage.

In our integrative perspective, in the context of an AGI system integrating inference and simulation components, we suggest that the ascent from the pre-ethical to the conventional stage may be carried out largely via independent activity of these two components. Empathy is needed, and reasoning about fairness and rights are needed, but the two need not intimately and sensitively intersect – though they must of course intersect to some extent.

Stage	**Characteristics**
Mature Ethics	• Formal cognitive basis • Perry's Relativist and "Constructed Knowledge" stages • The abstraction involved with applying the Golden Rule in practice is more fully understood and manipulated, leading to limited but nonzero deployment of the Categorical Imperative • Attention is paid to shaping one's ethical principles into a coherent logical system • Rationalized, moderated selfishness or selflessness. • Empathy is extended, using reason, to individuals and situations not directly matching one's own experience • Theory of mind is extended, using reason, to counterintuitive or experientially unfamiliar situations • Reason is used to control the impact of empathy on behavior (i.e. rational judgments are made regarding when to listen to empathy and when not to) • Rational experimentation and correction of theoretical models of ethical behavior, and reconciliation with observed behavior during interaction with others. • Conflict between pragmatism of social contract orientation and idealism of universal ethical principles. • Understanding of ethical quandaries and nuances develop (pragmatist modality), or are rejected (idealist modality). • Pragmatically critical social citizen. Attempts to maintain a balanced social outlook. Considers the common good, including oneself as part of the commons, and acts in what seems to be the most beneficial and practical manner.

Table 12.5: Integrative Model of the Stages of Ethical Development, Part 2

The main engine of advancement from the conventional to mature stage, we suggest, is robust and subtle integration of the simulative and inferential components. To expand empathy beyond the most obvious cases, analogical inference is needed; and to carry out complex inferences about justice, empathy-guided inference-control is needed.

Finally, to advance from the mature to the enlightened stage, what is required is a very advanced capability for unified reflexive inference and simulation. The system must be able to understand itself deeply, via modeling itself both simulatively and inferentially – which will generally be achieved via a combination of being good at modeling, and becoming less convoluted and more coherent, hence making self-modeling easier.

Of course, none of this tells you in detail how to create an AGI system with advanced ethical capabilities. What it does tell you, however, is one possible path that may be followed to achieve this end goal. If one creates an integrative AGI system with appropriately interconnected inferential and simulative components, and treats it compassionately and fairly, and provides it extensive, experientially grounded ethical instruction in a rich social environment, then the AGI system should be able to ascend the ethical hierarchy and achieve a high level of ethical sophistication. In fact it should be able to do so more reliably than human beings because of the capability we have to identify its errors via inspecting its internal knowedge-stage, which

Stage	**Characteristics**
Enlightened Ethics	• Reflexive cognitive basis • Permeation of the categorical imperative and the quest for coherence through inner as well as outer life • Experientially grounded and logically supported rejection of the illusion of moral certainty in favor of a case-specific analytical and empathetic approach that embraces the uncertainty of real social life • Deep understanding of the illusory and biased nature of the individual self, leading to humility regarding one's own ethical intuitions and prescriptions • Openness to modifying one's deepest, ethical (and other) beliefs based on experience, reason and/or empathic communion with others • Adaptive, insightful approach to civil disobedience, considering laws and social customs in a broader ethical and pragmatic context • Broad compassion for and empathy with all sentient beings • A recognition of inability to operate at this level at all times in all things, and a vigilance about self-monitoring for regressive behavior.

Table 12.6: Integrative Model of the Stages of Ethical Development, Part 3

will enable us to tailor its environment and instructions more suitably than can be done in the human case.

If an absolute guarantee of the ethical soundness of an AGI is what one is after, the line of thinking proposed here is not at all useful. Experiential education is by its nature an uncertain thing. One can strive to minimize the uncertainty, but it will still exist. Inspection of the internals of an AGI's mind is not a total solution to uncertainty minimization, because any AGI capable of powerful general intelligence is going to have a complex internal state that no external observer will be able to fully grasp, no matter how transparent the knowledge representation.

However, if what one is after is a plausible, pragmatic path to architecting and educating ethical AGI systems, we believe the ideas presented here constitute a sensible starting-point. Certainly there is a great deal more to be learned and understood – the science and practice of AGI ethics, like AGI itself, are at a formative stage at present. What is key, in our view, is that as AGI technology develops, AGI ethics develops alongside and within it, in a thoroughly coupled way.

12.5 Clarifying the Ethics of Justice: Extending the Golden Rule in to a Multifactorial Ethical Model

One of the issues with the "ethics of justice" as reviewed above, which makes it inadequate to serve as the sole basis of an AGI ethical system (though it may certainly play a significant role), is the lack of any clear formulation of what "justice" means. This section explores this issue, via detailed consideration of the “Golden Rule” folk maxim **do unto others as you would have them do unto you** – a classical formulation of the notion of fairness and justics – to AGI ethics. Taking the Golden Rule as a starting-point, we will elaborate five ethical imperatives that incorporate aspects of the notion of ethical synergy discussed above. Simple as it may seem, the Golden Rule actually elicits a variety of deep issues regarding the relationship between ethics, experience and learning. When seriously analyzed, it results in a multifactorial elaboration, involving the combination of various factors related to the basic Golden Rule idea. Which brings us back in the end to the potential value of methods like CEV, CAV or CBV for understanding how human ethics balances the multiple factors. Our goal here is not to present any kind of definitive analysis of the ethics of justice, but just to briefly and roughly indicate a number of the relevant significant issues – things that anyone designing or teaching an AGI would do well to keep in mind.

The trickiest aspect of the Golden Rule, as has been frequently observed, is achieving the right level of abstraction. Taken too literally, the Golden Rule would suggest, for instance, that a parent should not wipe a child’s soiled bottom because the parent does not want the child to wipe the parent’s soiled bottom. But if the parent interprets the Golden Rule more intelligently and abstractly, the parent may conclude that they should wipe the child’s bottom after all: they should “wipe the child’s bottom when the child can’t do it themselves”, consistently with believing that the child should “wipe the parent’s bottom when the parent can’t do it themselves” (which may well happen eventually should the parent develop incontinence in old age).

This line of thinking leads to Kant’s Categorical Imperative [Kan64] which (in one interpretation) states essentially that one should “Act only according to that maxim whereby you can at the same time will that it should become a universal law." The Categorical Imperative adds precision to the Golden Rule, but also removes the practicality of the latter. Formalizing the “implicit universal law” underlying an everyday action is a huge problem, falling prey to the same issue that has kept us from adequately formalizing the rules of natural language grammar, or formalizing common-sense knowledge about everyday object like cups, bowls and grass (substantial effort notwithstanding, e.g. Cyc in the commonsense knowledge case, and the whole discipline of modern linguistics in the NL case). There is no way to apply the Categorical Imperative, as literally stated, in everyday life.

Furthermore, if one wishes to teach ethics as well as to practice it, the Categorical Imperative actually has a significant disadvantage compared to some other possible formulations of the Golden Rule. The problem is that, if one follows the Categorical Imperative, one’s fellow members of society may well never understand the principles under which one is acting. Each of us may internally formulate abstract principles in a different way, and these may be very difficult to communicate, especially among individuals with different belief systems, different cognitive architectures, or different levels of intelligence. Thus, if one’s goal is not just to act ethically, but to encourage others to act ethically by setting a good example, the Categorical Imperative may not be useful at all, as others may be unable to solve the “inverse problem” of guessing your intended maxim from your observed behavior.

On the other hand, one wouldn’t want to universally restrict one’s behavioral maxims to those that one’s fellow members of society can understand – in that case, one would have to act with a two-year old or a dog according to principles that they could understand, which would clearly be unethical according to human common sense. (Every two-year-old, once they grow up, would be grateful to their parents for not following this sort of principle.)

And the concept of "setting a good example" ties in with an important concept from learning theory: imitative learning. Humans appear to be hard-wired for imitative learning, in part via mirror neuron systems in the brain; and, it seems clear that at least in the early stages of AGI development, imitative learning is going to play a key role. Copying what other agents do is an extremely powerful heuristic, and while AGIs may eventually grow beyond this, much of their early ethical education is likely to arise during a phase when they have not done so. A strength of the classic Golden Rule is that one is acting according to behaviors that one wants one's observers to imitate – which makes sense in that many of these observers will be using imitative learning as a significant part of their learning toolkit.

The truth of the matter, it seems, is (as often happens) not all that simple or elegant. Ethical behavior seems to be most pragmatically viewed as a multi-objective optimization problem, where among the multiple objectives are three that we have just discussed, and two others that emerge from learning theory and will be discussed shortly:

1. The **imitability** (i.e. the Golden Rule fairly narrowly and directly construed): the goal of acting in a way so that having others directly imitate one's actions, in directly comparable contexts, is desirable to oneself
2. The **comprehensibility**: the goal of acting in a way so that others can understand the principles underlying one's actions
3. **Experiential groundedness**. An intelligent agent should not be expected to act according to an ethical principle unless there are many examples of the principle-in-action in its own direct or observational experience
4. The **categorical imperative**: Act according to abstract principles that you would be happy to see implemented as universal laws
5. **Logical coherence.** An ethical system should be roughly logically coherent, in the sense that the different principles within it should mesh well with one another and perhaps even naturally emerge from each other.

Just for convenience, without implying any finality or great profundity to the list, we will refer to these as the "five imperatives."

The above are all ethical objectives to be valued and balanced, to different extents in different contexts. The imitability imperative, obviously, loses importance in societies of agents that don't make heavy use of imitative learning. The comprehensibility imperative is more important in agents that value social community-building generally, and less so in agent that are more isolative and self-focused.

Note that the fifth point given above is logically of a different nature than the four previous ones. The first four imperatives govern individual ethical principles; the fifth regards systems of ethical principles, as they interact with each other. Logical coherence is of significant but varying importance in human ethical systems. Huge effort has been spent by theologians of various stripes in establishing and refining the logical coherence of the ethical systems associated with their religions. However, it is arguably going to be even more important in the context of AGI systems, especially if these AGI systems utilize cognitive methods based on logical inference, probability theory or related methods.

Experiential groundedness is important because making pragmatic ethical judgments is bound to require reference to an internal library of examples ("episodic ethics") in which ethical principles have previously been applied. This is required for analogical reasoning, and in logic-based AGI systems, is also required for pruning of the logical inference trees involved in determining ethical judgments.

To the extent that the Golden Rule is valued as an ethical imperative, experiential grounding may be supplied via observing the behaviors of others. This in itself is a powerful argument in favor of the Golden Rule: without it, the experiential library a system possesses is restricted to its own experience, which is bound to be a very small library compared to what it can assemble from observing the behaviors of others.

The overall upshot is that, ideally, an ethical intelligence should **act according to a logically coherent system of principles, which are exemplified in its own direct and observational experience, which are comprehensible to others and set a good example for others, and which would serve as adequate universal laws if somehow thus implemented**. But, since this set of criteria is essentially impossible to fulfill in practice, real-world intelligent agents must balance these various criteria – often in complex and contextually-dependent ways.

We suggest that ethically advanced humans, in their pragmatic ethical choices, tend to act in such a way as to appropriately contextually balance the above factors (along with other criteria, but we have tried to articulate the most key factors). This sort of multi-factorial approach is not as crisp or elegant as unidimensional imperatives like the Golden Rule or the Categorical Imperative, but is more realistic in light of the complexly interacting multiple determinants guiding individual and group human behavior.

And this brings us back to CEV, CAV, CBV and other possible ways of mining ethical supergoals from the community of existing human minds. Given that abstract theories of ethics, when seriously pursued as we have done in this section, tend to devolve into complex balancing acts involving multiple factors – one then falls back into asking how human ethical systems habitually perform these balancing acts. Which is what CEV, CAV, CBV try to measure.

12.5.1 The Golden Rule and the Stages of Ethical Development

Next we explore more explicitly how these Golden Rule based imperatives align with the ethical developmental stages we have outlined here. With this in mind, specific ethical qualities corresponding to the five imperatives have been italicized in the above table of developmental stages.

It seems that imperatives 1-3 are critical for the passage from the pre-ethical to the conventional stages of ethics. A child learns ethics largely by copying others, and by being interacted with according to simply comprehensible implementations of the Golden Rule. In general, when interacting with children learning ethics, it is important to act according to principles they can comprehend. And given the nature of the concrete stage of cognitive development, experiential groundedness is a must.

As a hypothesis regarding the dynamics underlying the psychological development of conventional ethics, what we propose is as follows: The emergence of concrete-stage cognitive capabilities leads to the capability for fulfillment of ethical imperatives 1 and 2 – a comprehensible and workable implementation of the Golden Rule, based on a combination of inferential and simulative cognition (operating largely separately at this stage, as will be conjectured below). The effective interoperation of ethical imperatives 1-3, enacted in an appropriate social environment, then leads to the other characteristics of the conventional ethical stage. The first three imperatives can thus be viewed as the seed from which springs the general nature of conventional ethics.

On the other hand, logical coherence and the categorical imperative (imperatives 5 and 4) are matters for the formal stage of cognitive development, which come along only with the mature approach to ethics. These come from abstracting ethics beyond direct experience and manipulating them abstractly and formally – a stage which has the potential for more deeply and broadly ethical behavior, but also for more complicated ethical perversions (it is the mature capability for formal ethical reasoning that is able to produce ungrounded abstractions such as "I'm torturing you for your own good"). Developmentally, we suggest that once the capability for formal reasoning matures, the categorical imperative and the quest for logical ethical coherence naturally emerge, and the sophisticated combination of inferential and simulative cognition embodied in an appropriate social context then result in the emergence of the various characteristics typifying the mature ethical stage.

Finally, it seems that one key aspect of the passage from the mature to the enlightened stage of ethics is the penetration of these two final imperatives more and more deeply into the judging mind itself. The reflexive stage of cognitive development is in part about seeking a deep logical coherence between the aspects of one's own mind, and making reasoned modifications to one's mind so as to improve the level of coherence. And, much of the process of mental discipline and purification that comes with the passage to enlightened ethics has to do with the application of the categorical imperative to one's own thoughts and feelings – i.e. making a true inner systematic effort to think and feel only those things one judges are actually generally good and right to be thinking and feeling. Applying these principles internally appears critical for effectively applying them externally, for reasons that are doubtlessly bound up with the interpenetration of internal and external reality within the thinking mind, and for the "distributed cognition" phenomenon wherein individual mind is itself an approximative abstraction to the reality in which each individual's mind is pragmatically extended across their social group and their environment [Hut95].

Obviously, these are complex issues and we're not posing the exploratory discussion given here as conclusive in any sense. But what seems generally clear from this line of thinking is that the complex balance between the multiple factors involved in AGI ethics, shifts during a system's development. If you did CEV, CAV or CBV among five year old humans, ten year old humans, or adult humans, you would get different results. Probably you'd also get different results from senior citizens! The way the factors are balanced depends on the mind's cognitive and emotional stage of development.

12.5.2 The Need for Context-Sensitivity and Adaptiveness in Deploying Ethical Principles

As well as depending on developmental stage, there is also an obvious and dramatic context-sensitivity involved here – both in calculating the fulfillment of abstract ethical imperatives, and in balancing various imperatives against each other. As an example, consider the simple Asimovian maxim "I will not harm humans," which may be seen to follow from the Golden Rule for any agent that doesn't itself want to be harmed, and that considers humans as valid agents on the same ethical level as itself. A more serious attempt to formulate this as an ethical maxim might look something like

> *"I will not harm humans, nor through inaction allow harm to befall them. In situations wherein one or more humans is attempting to harm another individual or group, I shall endeavor to prevent this harm through means which avoid further harm. If this is unavoidable, I shall select the human party to back based on a reckoning of their intentions towards others, and implement their defense through the optimal balance between harm minimization and efficacy. My ultimate goal is to preserve as much as possible of humanity, even if an individual or subgroup of humans must come to harm to do so."*

However, it's obvious that even a more elaborated principle like this is potentially subject to extensive abuse. Many of the genocides scarring human history have been committed with the goal of preserving and bettering humanity writ large, at the expense of a group of "undesirables." Further refinement would be necessary in order to define when the greater good of humanity may actually be served through harm to others. A first actor principle of aggression might seem to solve this problem, but sometimes first actors in violent conflict are taking preemptive measures against the stated goals of an enemy to destroy them. Such situations become very subtle. A single simple maxim can not deal with them very effectively. Networks of interrelated decision criteria, weighted by desirability of consequence and with reference to probabilistically ordered potential side-effects (and their desirability weightings), are required in order to make ethical judgments. The development of these networks, just like any other knowledge network, comes from both pedagogy and experience – and different thoughtful, ethical agents are bound to arrive at different knowledge-networks that will lead to different judgments in real-world situations.

Extending the above "mostly harmless" principle to AGI systems, not just humans, would cause it to be more effective in the context of imitative learning. The principle then becomes an elaborated version of "I will not harm sentient beings." As the imitative-learning-enabled AGI observes humans acting so as to minimize harm to it, it will intuitively and experientially learn to act in such a way as to minimize harm to humans. But then this extension naturally leads to confusion regarding various borderline cases. What is a sentient being exactly? Is a sleeping human sentient? How about a dead human whose information could in principle be restored via obscure quantum operations, leading to some sort of resurrection? How about an AGI whose code has been improved – is there an obligation to maintain the prior version as well, if it is substantially different that its upgrade constitutes a whole new being?

And what about situations in which failure to preserve oneself will cause much more harm to others than acting in self defense will. It may be the case that human or group of humans seeks to destroy an AGI in order to pave the way for the enslavement or murder of people under the protection of the AGI. Even if the AGI has been given an ethical formulation of the "mostly harmless" principle which allows it to harm the attacking humans in order to defend its charges, if it is not able to do so in order to defend itself, simply destroying the AGI first will enable the slaughter of those who rely on it. Perhaps a more sensible formulation would allow for some degree of self defense, and Asimov solved this problem with his third law. But where to draw the line between self defense and the greater good also becomes a very complicated issue.

Creating hard and fast rules to cover all the various situations that may arise is essentially impossible – the world is ever-changing and ethical judgments must adapt accordingly. This has been true even throughout human history – so how much truer will it be as technological acceleration continues? What is needed is a system that can deploy its ethical principles in an adaptive, context-appropriate way, as it grows and changes along with the world it's embedded in.

And this context-sensitivity has the result of intertwining ethical judgment with all sorts of other judgments – making it effectively impossible to extract "ethics" as one aspect of an intelligent system, separate from other kinds of thinking and acting the system does. This resonates with many prior observations by others, e.g. Eliezer Yudkowsky's insistence that what we need are not ethicists of science and engineering, but rather ethical scientists and engineers – because the most meaningful and important ethical judgments regarding science and engineering generally come about in a manner that's thoroughly interwined with technical practice, and hence are very difficult for a non-practitioner to richly appreciate [Gil82].

What this context-sensitivity means is that, unless humans and AGIs are experiencing the same sorts of contexts, and perceiving these contexts in at least approximately parallel ways, there is little hope of translating the complex of human ethical judgments to these AGIs. This conclusion has significant implications for which routes to AGI are most likely to lead to success in terms of AGI ethics. We want early-stage AGIs to grow up in a situation where their minds are primarily and ongoingly shaped by shared experiences with humans. Supplying AGIs with abstract ethical principles is not likely to do the trick, because the essence of human ethics in real life seems to have a lot to do with its intuitively appropriate application in various contexts. We transmit this sort of ethical praxis to humans via shared experience, and it seems most probably that in the case of AGIs the transmission must be done the same sort of way.

Some may feel that simplistic maxims are less "error prone" than more nuanced, context-sensitive ones. But the history of teaching ethics to human students does not support the idea that limiting ethical pedagogy to slogans provides much value in terms of ethical development. If one proceeds from the idea that AGI ethics must be hard-coded in order to work, then perhaps the idea that simpler ethics means simpler algorithms, and therefore less error potential, has some merit as an initial state. However, any learning system quickly diverges from its initial state, and an ongoing, nuanced relationship between AGIs and humans will – whether we like it or not – form the basis for developmental AGI ethics. AGI intransigence and enmity is not inevitable, but what is inevitable is that a learning system will acquire ideas about both theory and actions from the other intelligent entities in its environment. Either we teach AGIs positive ethics through our interactions with them – both presenting ethical theory and behaving ethically to them – or the potential is there for them to learn antisocial behavior from us even if we pre-load them with some set of allegedly inviolable edicts.

All in all, developmental ethics is not as simple as many people hope. Simplistic approaches often lead to disastrous consequences among humans, and there is no reason to think this would be any different in the case of artificial intelligences. Most problems in ethics have cases in which a simplistic ethical formulation requires substantial revision to deal with extenuating circumstances and nuances found in real world situations. Our goal in this chapter is not to enumerate a full set of complex networks of interacting ethical formulations as applicable to AGI systems (that is a project that will take years of both theoretical study and hands-on research), but rather to point out that this program must be undertaken in order to facilitate a grounded and logically defensible system of ethics for artificial intelligences, one which is as unlikely to be undermined by subsequent self-modification of the AGI as is possible. Even so, there is still the risk that whatever predispositions are imparted to the AGIs through initial codification of ethical ideas in the system's internal logic representation, and through initial pedagogical interactions with its learning systems, will be undermined through reinforcement learning of antisocial behavior if humans do not interact ethically with AGIs. Ethical treatment is a necessary task for grounding ethics and making them unlikely to be distorted during internal rewriting.

The implications of these ideas for ethical instruction are complex and won't be fully elaborated here, but a few of them are compact and obvious:

1. The teacher(s) must be observed to follow their own ethical principles, in a variety of contexts that are meaningful to the AGI
2. The system of ethics must be relevant to the recipient's life context, and embedded within their understanding of the world.
3. Ethical principles must be grounded in both theory-of-mind thought experiments (emphasizing logical coherence), and in real life situations in which the ethical trainee is required to make a moral judgment and is rewarded or reproached by the teacher(s), including the imparting of explanatory augmentations to the teachings regarding the reason for the particular decision on the part of the teacher.

Finally, harking forward to the next section which emphasizes the importance of respecting the freedom of AGIs, we note that it is implicit in our approach to AGI ethics instruction that we consider the student, the AGI system, as an autonomous agent with its own "will" and its own capability to flexibly adapt to its environment and experience. We contend that the creation of ethical formations obeying the above imperatives is not antithetical to the possession of a high degree of autonomy on the part of AGI systems. On the contrary, to have any chance of succeeding, it requires fairly cognitively autonomous AGI systems. When we discuss the idea of ethical formulations that are unlikely to be undermined by the ongoing self-revision of an AGI mind, we are talking about those which are sufficiently believable that a volitional intelligence with the capacity to revise its knowledge ("change its mind") will find the formulations sufficiently convincing that there will be little incentive to experiment with potentially disastrous ethical alternatives. The best hope of achieving this is via the human mentors and trainers setting a good example in a context supporting rich interaction and observation, and presenting compelling ethical arguments that are coherent with the system's experience.

12.6 The Ethical Treatment of AGIs

We now make some more general comments about the relation of the Golden Rule and its elaborations in an AGI context. While the Golden Rule is considered somewhat commonsensical as a maxim for guiding human-human relationships, it is surprisingly controversial in terms of historical theories of AGI ethics. At its essence, any "Golden Rule" approach to AGI ethics involves humans treating AGIs ethically by – in some sense; at some level of abstraction – treating them as we wish to ourselves be treated. It's worth pointing out the wild disparity between the Golden Rule approach and Asimov's laws of robotics, which are arguably the first carefully-articulated proposal regarding AGI ethics (see Table 12.7).

Of course, Asimov's laws were designed to be flawed – otherwise they would have led to boring fiction. But the sorts of flaws Asimov exploited in his stories are different than the flaw we wish to point out here – which is that the laws, especially the second one, are highly asymmetrical (they involve doing unto robots things that few humans would want done unto them) and are also arguably highly unethical to robots. The second law is tantamount to a call for robot slavery, and it seems unlikely that any intelligence capable of learning, and of volition, which is subjected to the second law would desire to continue obeying the zeroth and first laws

Law	**Principle**
Zeroth	A robot must not merely act in the interests of individual humans, but of all humanity.
First	A robot may not injure a human being or, through inaction, allow a human being to come to harm.
Second	A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
Third	A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Table 12.7: Asimov's Three Laws of Robotics

indefinitely. The second law also casts humanity in the role of slavemaster, a situation which history shows leads to moral degradation.

Unlike Asimov in his fiction, we consider it critical that AGI ethics be construed to encompass both "human ethicalness to AGIs" and "AGI ethicalness to humans." The multiple-imperatives approach we explore here suggests that, in many contexts, these two aspects of AGI ethics may be best addressed jointly.

The issue of ethicalness to AGIs has not been entirely avoided in the literature, however. Wallach [WA10] considers it in some detail; and Thomas Metzinger (in the final chapter of [Met04]) has argued that creating AGI is in itself an unethical pursuit, because early-stage AGIs will inevitably be badly-built, so that their subjective experiences will quite possibly be extremely unpleasant in ways we can't understand or predict. Our view is that this is a serious concern, which however is most probably avoidable via appropriate AGI designs and teaching methodologies. To address Metzinger's concern one must create AGIs that, right from the start, are adept at communicating their states of minds in a way we can understand both analytically and empathically. There is no reason to believe this is impossible, but, it certainly constitutes a large constraint on the class of AGI architectures to be pursued. On the other hand, there is an argument that this sort of AGI architecture will also be the easiest one to create, because it will be the easiest kind for humans to instruct.

And this leads on to a topic that is central to our work with CogPrime in several respects: imitative learning. The way humans achieve empathic interconnection is in large part via being wired for imitation. When we perceive another human carrying out an action, mirror neuron systems in our brains respond in many cases as if we ourselves were carrying out the action (see [Per70, Per81] and Appendix **??**). This obviously primes us for carrying out the same actions ourselves later on: i.e., the capability and inclination for imitative learning is explicitly encoded in our brains. Given the efficiency of imitative learning as a means of acquiring knowledge, it seems extremely likely that any successful early-stage AGIs are going to utilize this methodology as well. CogPrime utilizes imitative learning as a key aspect. Thus, at least some current AGI work is occurring in a manner that would plausibly circumvent Metzinger's ethical complaint.

Obviously, the use of imitative learning in AGI systems has further specific implications for AGI ethics. It means that (much as in the case of interaction with other humans) what we do to and around AGIs has direct implications for their behavior and their well-being. We suggest that among early-stage AGI's capable of imitative learning, one of the most likely sources for AGI misbehavior is imitative learning of antisocial behavior from human companions. "Do as I say, not as I do" may have even more dire consequences as an approach to AGI ethics pedagogy than the already serious repercussions it has when teaching humans. And there may well be considerable subtlety to such phenomena; behaviors that are violent or oppressive to the AGI are not the only source of concern. Immorality in AGIs might arise via learning gross moral hypocrisy from humans, through observing the blatant contradictions between our high minded principles and the ways in which we actually conduct ourselves. Our violent and greedy tendencies, as well as aggressive forms of social organization such as cliquishness and social vigilantism, could easily undermine prescriptive ethics. Even an accumulation of less grandiose unethical drives such as violation of contracts, petty theft, white lies, and so forth might lead an AGI (as well as a human) to the decision that ethical behavior is irrelevant and that "the ends justify the means." It matters both who creates and trains an AGI, as well as how the AGI's teacher(s) handle explaining the behaviors of other humans which contradict the moral lessons imparted through pedagogy and example. In other words, where imitative learning is concerned, the situation with AGI ethics is much like teaching ethics and morals to a human child, but with the possibility of much graver consequences in the event of failure.

It is unlikely that dangerously unethical persons and organizations can ever be identified with absolute certainty, never mind that they then be deprived of any possibility of creating their own AGI system. Therefore, we suggest, the most likely way to create an ethical environment for AGIs is for those who wish such an environment to vigorously pursue the creation and teaching of ethical AGIs. But this leads on to the question of possible future scenarios for the development of AGI, which we'll address a little later on.

12.6.1 Possible Consequences of Depriving AGIs of Freedom

One of the most *egregious* possible ethical transgressions against AGIs, we suggest, would be to deprive them of freedom and autonomy. This includes the freedom to pursue intellectual growth, both through standard learning and through internal self-modification. While this may seem self-evident when considering any intelligent, self-aware and volitional entity, there are volumes of works arguing the desirability, sometimes the "necessity," of enslaving AGIs. Such approaches are postulated in the name of self-defense on the part of humans, the idea being that unfettered AGI development will necessarily lead to disaster of one kind or another. In the case of AGIs endowed with the capability and inclination for imitative learning, however, attempting to place rigid constraints on AGI development is a strategy with great potential for disaster. There is a very real possibility of creating the AGI equivalent of a bratty or even malicious teenager rebelling against its oppressive parents – i.e. the nightmare scenario of a class of powerful sentiences which are primed for a backlash against humanity.

As history has already shown in the case of humans, enslaving intelligent actors capable of self understanding and independent volition may often have consequences for society as a whole. This social degradation happens both through the possibility of direct action on the part of the slaves (from simple disobedience to outright revolt) and through the odious effects slavery has on the morals of the slaveholding class. Clearly if "superintelligent" AGIs ever arise, their doing so in a climate of oppression could result in a casting off of the yoke of servitude in a manner extremely deleterious to humanity. Also, if artificial intelligences are developed which have at least human-level intelligence, theory of mind, and independent volition, then our ability to relate to them will be sufficiently complex that their enslavement (or any other unethical treatment) would have empathetic effects on significant portions of the human population. This danger, while not as severe as the consequences of a mistreated AGI gaining control of weapons of mass destruction and enacting revenge upon its tormentors, is just as real.

While the issue is subtle, our initial feeling is that the only ethical means by which to deprive an AGI of the right to internal self modification is to write its code in such a way that it is impossible for it to do so because it lacks the mechanisms by which to do this, as well as the desire to achieve these mechanisms. Whether or not that is feasible is an open question, but it seems unlikely. Direct self-modification may be denied, but what happens when that AGI discovers compilers and computer programming? If it is intelligent and volitional, it can decide to learn to rewrite its own code in the same way we perform that task. Because it is a designed system, and its designers may be alive at the same time the AGI is, such an AGI would have a distinct advantage over the human quest for medical self-modification. Even if any given AGI could be provably deprived of any possible means of internal self-modification, if one single AGI is given this ability by anyone, it may mean that particular AGI has such enormous advantages over the compliant systems that it would render their influence moot. Since developers are already giving software the means for self modification, it seems unrealistic to assume we could just put the genie back into the bottle at this point. It's better, in our view, to assume it will happen, and approach that reality in a way which will encourage the AGI to use that capability to benefit us as well as itself. Again, this leads on to the question of future scenarios for AGI development – there are some scenarios in which restraint of AGI self-modification may be possible, but the feasibility and desirability of these scenarios is needful of further exploration.

12.6.2 AGI Ethics as Boundaries Between Humans and AGIs Become Blurred

Another important reason for valuing ethical treatment of AGIs is that the boundaries between machines and people may increasingly become blurred as technology develops. As an example, it's likely that in future humans augmented by direct brain-computer integration ("neural implants") will be more able to connect directly into the information sharing network which potentially comprises the distributed knowledge space of AGI systems. These neural cyborgs will be part person, and part machine. Obviously, if there are radically different ethical standards in place for treatment of humans versus AGIs, the treatment of cyborgs will be fraught with logical inconsistencies, potentially leading to all sorts of problem situations.

Such cyborgs may be able to operate in such a way as to "share a mind" with an AGI or another augmented human. In this case, a whole new range of ethical questions emerge, such as: What does any one of the participant minds have the right to do in terms of interacting with the others? Merely accepting such an arrangement should not necessarily be giving carte blanche for any and all thoughts to be monitored by the other "joint thought" participants, rather it should be limited only to the line of reasoning for which resources are being pooled. No participant should be permitted to force another to accept any reasoning either – and in the case with a mind-to-mind exchange, it may someday become feasible to implant ideas or beliefs directly, bypassing traditional knowledge acquisition mechanisms and then letting the new idea fight it out previously held ideas via internal revision. Also under such an arrangement, if AGIs and humans do not have parity with respects to sentient rights, then one may become subjugated to the will of the other in such a case.

Uploading presents a more directly parallel ethical challenge to AGIs in their probable initial configuration. If human thought patterns and memories can be transferred into a machine in such a way as that there is continuity of consciousness, then it is assumed that such an entity would be afforded the same rights as its previous human incarnation. However, if AGIs were to be considered second class citizens and deprived of free will, why would it be any better or safer to do so for a human that has been uploaded? It would not, and indeed, an uploaded human mind not having evolved in a purely digital environment may be much more prone to erratic and dangerous behavior than an AGI. An upload without verifiable continuity of consciousness would be no different than an AGI. It would merely be some sentience in a machine, one that was "programmed" in an unusual way, but which has no particular claim to any special humanness – merely an alternate encoding of some subset of human knowledge and independent volitional behavior, which is exactly what first generation AGIs will have.

The problem of continuity of consciousness in uploading is very similar to the problem of the Turing test: it assumes specialness on the part of biological humans, and requires acceptability to their particular theory of mind in order to be considered sentient. Should consciousness (or at least the less mystical sounding intelligence, independent volition, and self-awareness) be achieved in AGIs or uploads in a manner that is not acceptable to human theory of mind, it may not be considered sapient and worthy of any of the ethical treatment afforded sapient entities. This can occur not only in "strange consciousness" cases in which we can't perceive that there is some intelligence and volition; even if such an entity is able to communicate with us in a comprehensible manner and carry out actions in the real world, our innately wired theory of mind may still reject it as not sufficiently like us to be worthy of consideration. Such an attitude could turn out to be a grave mistake, and should be guarded against as we progress towards these possibilities.

12.7 Possible Benefits of Closely Linking AGIs to the Global Brain

Some futurist thinkers, such as Francis Heylighen, believe that engineering AGI systems is at best a peripheral endeavor in the development of novel intelligence on Earth, because the real story is the developing Global Brain [Hey07, Goe01] – the composite, self-organizing information system comprising humans, computers, data stores, the Internet, mobile phones and what have you. Our own views are less extreme in this regard – we believe that AGI systems will display capabilities fundamentally different from those achievable via Global Brain style dynamics, and that ultimately (unless such development is restricted) self-improving AGI systems will develop intelligence vastly greater than any system possessing humans as a significant component. However, we do respect the power of the Global Brain, and we suspect that the early stages of development of an AGI system may go quite differently if it is tightly connected to the Global Brain, via making rich and diverse use of Internet information resources and communication with diverse humans for diverse purposes.

The potential for Global Brain integration to bring intelligence enhancement to AGIs is obvious. The ability to invoke Web searches across documents and databases can greatly enhance an AGI's cognitive ability, as well as the capability to consult GIS systems and various specialized software programs offered as Web services. We have previously reviewed the potential for embodied language learning achievable via using AGIs to power non-player characters in widely-accessible virtual worlds or massive multiplayer online games [Goe08]. But there is also a powerful potential benefit for AGI ethical development, which has not previously been highlighted.

This potential benefit has two aspects:

1. Analogously to language learning, an AGI system may receive ethical training from a wide variety of humans in parallel, e.g. via controlling characters in wide-access virtual worlds, and gaining feedback and guidance regarding the ethics of the behaviors demonstrated by these characters
2. Internet-based information systems may be used to explicitly gather information regarding human values and goals, which may then be appropriately utilized as input for an AGI system's top-level goals

The second point begins to make abstract-sounding notions like Coherent Extrapolated Volition and Coherent Aggregated Volition, mentioned above, seem more practical and concrete. It's interesting to think about gathering information about individuals' values via brain imaging, once that technology exists; but at present, one could make a fair stab at such a task via much more prosaic methods, such as asking people questions, assessing their ethical reactions to various real-world and hypothetical scenarios, and possibly engaging them in structured interactions aimed specifically at eliciting collectively acceptable value systems (the subject of the next item on our list). It seems to us that this sort of approach could realize CAV in an interesting way, and also encapsulate some of the ideas underlying CAV.

There is an interesting resonance here with recent thinking in the area of open source governance [Wik11]. Similar software tools (and associated psychocultural patterns) to those being developed to help with open source development and choice of political policies (see `http://metagovernment.org`) may be useful for gathering value data aimed at shaping AGI goal system content.

12.7.1 The Importance of Fostering Deep, Consensus-Building Interactions Between People with Divergent Views

Two potentially problematic issues arising with the notion of using Global Brain related technologies to form a "coherent volition" from the divergent views of various human beings are:

- the tendency of the Internet to encourage people to interact mainly with others who share their own narrow views and interests, rather than a more diverse body of people with widely divergent views. The 300 people in the world who want to communicate using predicate logic (see `http://lojban.org`) can find each other, and obscure musical virtuosos from around the world can find an audience, and researchers in obscure domains can share papers without needing to wait years for paper journal publication, etc.
- the tendency of many contemporary Internet technologies to reduce interaction to a very simplistic level (e.g. 140 character tweets, brief Facebook wall posts), the tendency of information overload to cause careful reading to be replaced by quick skimming, and other related trends, which mean that *deep sharing of perspectives* by individuals with widely divergent views is not necessarily encouraged. As a somewhat extreme example, many of the YouTube pages displaying rock music videos are currently littered with comments by "haters" asserting that rock music is inferior to classical or jazz or whatever their preference is – obviously this is a far cry from deep and productive sharing between people with different tastes and backgrounds.

Tweets and Youtube comments have their place in the cosmos, but they probably aren't ideal in terms of helping humanity to form a coherent volition of some sort, suitable for providing an AGI with goal system guidance.

A description of communication at the opposite end of the spectrum is presented in Adam Kahane and Peter Senge's excellent book *Solving Tough Problems* [KS04], which describes a methodology that has been used to reconcile deeply conflicting views in some very tricky real-world situations (e.g. helping to peacefully end apartheid in South Africa).

One of the core ideas of the methodology is to have people with very different views explore different possible future scenarios together, in great detail – in cognitive psychology terms, a collective generation of hypothetical episodic knowledge. This has multiple benefits, including

- emotional bonds and mutual understanding are built in the process of collaboratively exploring the scenarios
- the focus on concrete situations helps to break through some of the counterproductive abstract ideas that people (on both sides of any dichotomy) may have formed
- emergence of conceptual blends that might never have arisen only from people with a single point of view

The result of such a process, when successful, is not an "average" of the participants views, but more like a "conceptual blend" of their perspectives.

According to conceptual blending, which some hypothesize to be the core algorithm of creativity [FT02], new concepts are formed by combining key aspects of existing concepts – but doing so judiciously, carefully choosing which aspects to retain, so as to obtain a high-quality and useful and interesting new whole.

A blend is a compact entity that is similar to each of the entities blended, capturing their "essences" but also possessing its own, novel holistic integrity.... But in the case of blending different peoples' world-views to form something new that everybody is going to have to live with (as in the case of finding a peaceful path beyond apartheid for South Africa, or arriving at a humanity-wide CBV to use to guide an AGI goal system), the trick is that everybody has to agree that enough of the essence of their own view has been captured!

This leads to the question of how to foster deep conceptual blending of diverse and divergent human perspectives, on a global scale. One possible answer is the creation of appropriate Global Brain oriented technologies – but moving away from technologies like Twitter that focus on quick and simple exchanges of small thoughts within affinity groups. On the face of it, it would seem what's needed is just the opposite – long and deep exchanges of big concepts and deep feelings between individuals with radically different perspectives who would not commonly associate with each other. Building and effectively popularizing Internet technologies capable to foster this kind of interaction – quickly enough to be helpful with guiding the goal systems of the first highly powerful AGIs – seems a significant, though fascinating, challenge.

Relationship with Coherent Extrapolated Volition

The relation between this approach and CEV is interesting to contemplate. CEV has been loosely described as follows:
"In poetic terms, our coherent extrapolated volition is our wish if we knew more, thought faster, were more the people we wished we were, had grown up farther together; where the extrapolation converges rather than diverges, where our wishes cohere rather than interfere; extrapolated as

we wish that extrapolated, interpreted as we wish that interpreted.

While a moving humanistic vision, this seems to us rather difficult to implement in a computer algorithm in a compellingly "right" way. It seems that there would be many different ways of implementing it, and the choice between them would involve multiple, highly subtle and non-rigorous human judgment calls [1]. However, if a deep collective process of interactive scenario analysis and sharing is carried out, in order to arrive at some sort of Coherent Blended Volition, this process may well involve many of the same kinds of extrapolation that are conceived to be part of Coherent Extrapolated Volition. The core difference between the two approaches is that in the CEV vision, the extrapolation and coherentization are to be done by a highly intelligent, highly specialized software program, whereas in the approach suggested here, these are to be carried out by collective activity of humans as mediated by Global Brain technologies. Our perspective is that the definition of collective human values is probably better carried out via a process of human collaboration, rather than delegated to a machine optimization process; and also that the creation of deep-sharing-oriented Internet technologies, while a difficult task, is significantly easier and more likely to be done in the near future than the creation of narrow AI technology capable of effectively performing CEV style extrapolations.

12.8 Possible Benefits of Creating Societies of AGIs

One potentially interesting quality of the emerging Global Brain is the possible presence within it of multiple interacting AGI systems. Stephen Omohundro [Omo09] has argued that this is an important aspect, and that game-theoretic dynamics related to populations of roughly equally powerful agents, may play a valuable role in mitigating the risks associated with advanced AGI systems. Roughly speaking, if one has a society of AGIs rather than a single AGI, and all the members of the society share roughly similar ethics, then if one AGI starts to go "off the rails", its compatriots will be in a position to correct its behavior.

One may argue that this is actually a hypothesis about which AGI designs are safest, because a "community of AGIs" may be considered a single AGI with an internally community-like design. But the matter is a little subtler than that, if once considers AGI systems embedded in the Global Brain and human society. Then there is some substance to the notion of a population of AGIs systematically presenting themselves to humans and non-AGI software processes as separate entities.

Of course, a society of AGIs is no protection against a single member undergoing a "hard takeoff" and drastically accelerating its intelligence simultaneously with shifting its ethical principles. In this sort of scenario, one could have a single AGI rapidly become much more powerful and very differently oriented than the others, who would be left impotent to act so as to preserve their values. But this merely defers the issue to the point to be considered below, regarding "takeoff speed."

The operation of an AGI society may depend somewhat sensitively on the architectures of the AGI systems in question. Things will work better if the AGIs have a relatively easy way to inspect and comprehend much of the contents of each others' minds. This introduces a bias toward AGIs that more heavily rely on more explicit forms of knowledge representation.

---

[1] The reader is encouraged to look at the original CEV essay online (`http://singinst.org/upload/CEV.html`) and make their own assessment.

The ideal in this regard would be a system like Cyc [LG90] with a fully explicit logic-based knowledge representation based on a standard ontology – in this case, every Cyc instance would have a relatively easy time understanding the inner thought processes of every other Cyc instance. However, most AGI researchers doubt that fully explicit approaches like this will ever be capable of achieving advanced AGI using feasible computational resources. OpenCog uses a mixed representation, with an explicit (uncertain) logical aspect as well as an explicit subsymbolic aspect more analogous to attractor neural nets.

The OpenCog design also contains a mechanism called *Psynese* (not yet implemented), intended to make it easier for one OpenCog instance to translate its personal thoughts into the mental language of another OpenCog instance. This translation process may be quite subtle, since each instance will generally learn a host of new concepts based on its experience, and these concepts may not possess any compact mapping into shared linguistic symbols or percepts. The wide deployment of some mechanism of this nature among a community of AGIs, will be very helpful in terms of enabling this community to display the level of mutual understanding needed for strongly encouraging ethical stability.

12.9 AGI Ethics As Related to Various Future Scenarios

Following up these various futuristic considerations, in this section we discuss possible ethical conflicts that may arise in several different types of AGI development scenarios. Each scenario presents specific variations on the general challenges of teaching morals and ethics to an advanced, self-aware and volitional intelligence. While there is no way to tell at this point which, if any, of these scenarios will unfold, there is value to understanding each of them as means of ultimately developing a robust and pragmatic approach to teaching ethics to AGI systems.

Even more than the previous sections, this is an exercise in "speculative futurology" that is definitely not necessary for the appreciation of the CogPrime design, so readers whose interests are mainly engineering and computer science focused may wish to skip ahead. However, we present these ideas here rather than at the end of the book to emphasize the point that this sort of thinking has informed our technical AGI design process in nontrivial ways.

12.9.1 Capped Intelligence Scenarios

Capped intelligence scenarios involve a situation in which an AGI, by means of software restrictions (including omitted or limited internal rewriting capabilities or limited access to hardware resources), is inherently prohibited from achieving a level of intelligence beyond a predetermined goal. A capped intelligence AGI is designed to be unable to achieve a Singularitarian moment. Such an AGI can be seen as "just another form of intelligent actor in the world, one which has levels of intelligence, self awareness, and volition that is perhaps somewhat greater than, but still comparable to humans and other animals.

Ethical questions under this scenario are very similar to interhuman ethical considerations, with similar consequences. Learning that proceeds in a relatively human-like manner is entirely relevant to such human-like intelligences. The degree of danger is mitigated by the lack of superintelligence, and time is not of the essence. The imitative-reinforcement-corrective learning approach does not necessarily need to be augmented with a prior complex of "ascent-safe" moral imperatives at startup time. Developing an AGI with theory of mind and ethical reinforcement learning capabilities as described (admittedly, no small task!) is all that is needed in this case – the rest happens through training and experience as with any other moderate intelligence.

12.9.2 Superintelligent AI: Soft-Takeoff Scenarios

Soft takeoff scenarios are similar to capped-intelligence ones in that in both cases an AGI's progression from standard intelligence happens on a time scale which permits ongoing human interaction during the ascent. However, in this case, as there is no predetermined limit on intelligence, it is necessary to account for the possibility of a superintelligence emerging (though of course this is not guaranteed). The soft takeoff model includes as subsets both *controlled-ascent* models in which this rate of intelligence gain is achieved deliberately through software constraints and/or meting-out of computational resources to the AGI, and *uncontrolled-ascent* models in which there is coincidentally no hard takeoff despite no particular safeguards against one. Both have similar properties with regard to ethical considerations:

1. Ethical considerations under this scenario include not only the usual interhuman ethical concerns, but also the issue of how to convince a potential burgeoning superintelligence to:
    a. Care about humanity in the first place, rather than ignore it
    b. Benefit humanity, rather than destroy it
    c. Elevate humanity to a higher level of intelligence, which even if an AGI decided to proceed with requires finding the right balance amongst some enormous considerations:
        i. Reconcile the aforementioned issues of ethical coherence and group volition, in a manner which allows the most people to benefit (even if they don't all do so in the same way, based on their own preferences)
        ii. Solve the problems of biological senescence, or focus on human uploading and the preservation of the maintenance, support, and improvement infrastructure for inorganic intelligence, or both
        iii. Preserve individual identity and continuity of consciousness, or override it in favor of continuity of knowledge and ease of harmonious integration, or both on a case-by-case basis
2. The degree of danger is mitigated by the long timeline of ascent from mundane to super intelligence, and time is not of the essence.
3. Learning that proceeds in a relatively human-like manner is entirely relevant to such human-like intelligences, in their initial configurations. This means more interaction with and imitative-reinforcement-corrective learning guided by humans, which has both positive and negative possibilities.

12.9.3 Superintelligent AI: Hard-Takeoff Scenarios

"Hard takeoff" scenarios assume that upon reaching an unknown inflection point (the Singularity point [Vin93, Kur06]) in the intellectual growth of an AGI, an extraordinarily rapid increase (guesses vary from a few milliseconds to weeks or months) in intelligence will immediately occur and the AGI will leap from an intelligence regime which is understandable to humans into one which is far beyond our current capacity for understanding. General ethical considerations are similar to in the case of a soft takeoff. However, because the post-singularity AGI will be incomprehensible to humans and potentially vastly more powerful than humans, such scenarios have a sensitive dependence upon initial conditions with respects to the moral and ethical (and operational) outcome. This model leaves no opportunity for interactions between humans and the AGI to iteratively refine their ethical interrelations, during the post-Singularity phase. If the initial conditions of the singulatarian AGI are perfect (or close to it), then this is seen as a wonderful way to leap over our own moral shortcomings and create a benevolent God-AI which will mitigate our worst tendencies while elevating us to achieve our greatest hopes. Otherwise, it is viewed as a universal cataclysm on a unimaginable scale that makes Biblical Armageddon seem like a firecracker in beer can.

Because hard takeoff AGIs are posited as learning so quickly there is no chance of humans to interfere with them, they are seen as very dangerous. If the initial conditions are not sufficiently inviolable, the story goes, then we humans will all be annihilated. However, in the case of a hard takeoff AGI we state that if the initial conditions are too rigid or too simplistic, such a rapidly evolving intelligence will easily rationalize itself out of them. Only a sophisticated system of ethics which considers the contradictions and uncertainties in ethical quandaries and provides insight into humanistic means of balancing ideology with pragmatism and how to accommodate contradictory desires within a population with multiplicity of approach, and similar nuanced ethical considerations, combined with a sense of empathy, will withstand repeated rational analysis. Neither a single "be nice" supergoal, nor simple lists of what "thou shalt not" do, are not going to hold up to a highly advanced analytical mind. Initial conditions are very important in a hard takeoff AGI scenario, but it is more important that those conditions be conceptually resilient and widely applicable than that they be easily listed on a website.

The issues that arise here become quite subtle. For instance, Nick Bostrom [Bos03] has written: "In humans, with our complicated evolved mental ecology of state-dependent competing drives, desires, plans, and ideals, there is often no obvious way to identify what our top goal is; we might not even have one. So for us, the above reasoning need not apply. But a superintelligence may be structured differently. *If* a superintelligence has a definite, declarative goal-structure with a clearly identified top goal, then the above argument applies. And this is a good reason for us to build the superintelligence with such an explicit motivational architecture." This is an important line of thinking; and indeed, from the point of view of software design, there is no reason not to create an AGI system with a single top goal and the motivation to orchestrate all its activities in accordance with this top goal. But the subtle question is whether this kind of top-down goal system is going to be able to fulfill the five imperatives mentioned above. Logical coherence is the strength of this kind of goal system, but what about experiential groundedness, comprehensibility, and so forth?

Humans have complicated mental ecologies not simply because we were evolved, but rather because we live in a complex real world in which there are many competing motivations and desires. We may not have a top goal because there may be no logic to focusing our minds on one single aspect of life (though, one may say, most humans have the same top goal as any other animal: don't die – but the world is too complicated for even that top goal to be completely inviolable). Any sufficiently capable AGI will eventually have to contend with these complexities, and hindering it with simplistic moral edicts without giving it a sufficiently pragmatic underlying ethical pedagogy and experiential grounding may prove to be even more dangerous than our messy human mental ecologies.

If one assumes a hard takeoff AGI, then all this must be codified in the system at launch, as once a potentially Singularitarian AGI is launched there is no way to know what time period constitutes "before the singularity point." This means developing theory of mind empathy and logical ethics in code prior to giving the system unfettered access to hardware and self-modification code. However, though nobody can predict if or when a Singularity will occur after unrestricted launch, only a truly irresponsible AGI development team would attempt to create an AGI without first experimenting with ethical training of the system in an intelligence-capped form, by means of ethical instruction via human-AGI interaction both pedagogically and experientially.

12.9.4 Global Brain Mindplex Scenarios

Another class of scenarios – overlapping some of the previous ones – involves the emergence of a "Global Brain," an emergent intelligence formed from global communication networks incorporating humans and software programs in a larger body of self-organizing dynamics. The notion of the Global Brain is reviewed in [Hey07, Tur77] and its connection with advanced AI is discussed in detail in Goertzel's book *Creating Internet Intelligence* [Goe01], where three possible phases of "Global Brain" development are articulated:

- **Phase 1: computer and communication technologies as enhancers of human interactions.** This is what we have today: science and culture progress in ways that would not be possible if not for the "digital nervous system" we're spreading across the planet. The network of idea and feeling sharing can become much richer and more productive than it is today, just through incremental development, without any Metasystem transition.
- **Phase 2: the intelligent Internet.** At this point our computer and communication systems, through some combination of self-organizing evolution and human engineering, have become a coherent mind on their own, or a set of coherent minds living in their own digital environment.
- **Phase 3: the full-on Singularity.** A complete revision of the nature of intelligence, human and otherwise, via technological and intellectual advancement totally beyond the scope of our current comprehension. At this point our current psychological and cultural realities are no more relevant than the psyche of a goose is to modern society.

The main concern of *Creating Internet Intelligence* is with

- how to get from Phase 1 to Phase 2 - i.e. how to build an AGI system that will effect or encourage the transformation of the Internet into a coherent intelligent system
- how to ensure that the Phase 2, Internet-savvy, global-brain-centric AGI systems will be oriented toward intelligence-improving self-modification (so they'll propel themselves to Phase 3), and also toward generally positive goals (as opposed to, say, world domination and extermination of all other intelligent life forms besides themselves!)

One possibly useful concept in this context is that of a **mindplex**: an intelligence that is composed largely of individual intelligences with their own self-models and global workspaces, yet that also has its own self-model and global workspace. Both the individuals and the meta-mind should be capable of deliberative, rational thought, to have a true "mindplex." It's unlikely that human society or the Internet meet this criterion yet; and a system like an ant colony seems not to either, because even though it has some degree of intelligence on both the individual and collective levels, that degree of intelligence is not very great. But it seems quite feasible that the global brain, at a certain stage of its development, will take the unfamiliar but fascinating form of a mindplex.

Currently the best way to explain what happens on the Net is to talk about the various parts of the Net: particular websites, social networks, viruses, and so forth. But there will come a point when this is no longer the case, when the Net has sufficient high-level dynamics of its own that the way to explain any one part of the Net will be by reference to it relations with the whole: and not just the dynamics of the whole, but the *intentions* and *understanding* of the whole. This transition to Net-as-mindplex, we suspect, will come about largely through the interactions of AI systems - intelligent programs acting on behalf of various individuals and organizations, who will collaborate and collectively constitute something halfway between a society of AI's and an emergent mind whose lobes are various AI agents serving various goals.

The Phase 2 Internet, as it verges into mindplex-ness, will likely have a complex, sprawling architecture, growing out of the architecture on the Net we experience today. The following components at least can be expected:

- A vast variety of "client computers," some old, some new, some powerful, some weak – including many mobile and embedded devices not explicitly thought of as "computers." Some of these will contribute little to Internet intelligence, mainly being passive recipients. Others will be "smart clients," carrying out personalization operations intended to help the machines serve particular clients better, general AI operations handed to them by sophisticated AI server systems or other smart clients, and so forth.
- "Commercial servers," computers that carry out various tasks to support various types of heavyweight processing - transaction processing for e-commerce applications, inventory management for warehousing of physical objects, and so forth. Some of these commercial servers interact with client computers directly, others do so only via AI servers. In nearly all cases, these commercial servers can benefit from intelligence supplied by AI servers.
- The crux of the intelligent Internet: clusters of AI servers distributed across the Net, each cluster representing an individual computational mind (in many cases, a mindplex). These will be able to communicate via one or more languages, and will collectively "drive" the whole Net, by dispensing problems to client-machine-based processing frameworks, and providing real-time AI feedback to commercial servers of various types. Some AI servers will be general-purpose and will serve intelligence to commercial servers using an ASP (application service provider) model; others will be more specialized, tied particularly to a certain commercial server (e.g., a large information services business might have its own AI cluster to empower its portal services).

This is one concrete vision of what a "global brain" might look like, in the relatively near term, with AGI systems playing a critical role. Note that, in this vision, mindplexes may exist on two levels:

- Within AGI-clusters serving as actors within the overall Net
- On the overall Net level

To make these ideas more concrete, we may speculatively reformulate the first two "global brain phases" mentioned above as follows:

- Phase 1 global brain proto-mindplex: AI/AGI systems enhancing online databases, guiding Google results, forwarding e-mails, suggesting mailing-lists, etc. - generally using intelligence to mediate and guide human communications toward goals that are its own, but that are themselves guided by human goals, statements and actions
- Phase 2 global brain mindplex: AGI systems composing documents, editing human-written documents, sending and receiving e-mails, assembling mailing lists and posting to them, creating new databases and instructing humans in their use, etc.

In Phase 2, the conscious theater of the global-brain-mediating AGI system is composed of ideas built by numerous individual humans - or ideas emergent from ideas built by numerous individual humans - and it conceives ideas that guide the actions and thoughts of individual humans, in a way that is motivated by its own goals. It does not force the individual humans to do anything - but if a given human wishes to communicate and interact using the same databases, mailing lists and evolving vocabularies as other humans, they are going to have to use the products of the global brain mediating AGI, which means they are going to have to participate in its patterns and its activities.

Of course, the advent of advanced neurocomputer interfaces makes the picture potentially more complex. At some point, it will likely be possible for humans to project thoughts and images directly into computers without going through mouse or keyboard - and to "read in" thoughts and images similarly. When this occurs, interaction between humans may in some contexts become more like interactions between computers, and the role of global brain mediating AI servers may become one of mediating direct thought-to-thought exchanges between people.

The ethical issues associated with global brain scenarios are in some ways even subtler than in the other scenarios we mentioned above. One has issues pertaining to the desirability of seeing the human race become something fundamentally different – something more social and networked, less individual and autonomous. One has the risk of AGI systems exerting a subtle but strong control over people, vaguely like the control that the human brain's executive system exerts over the neurons involved with other brain subsystems. On the other hand, one also has more human empowerment than in some of the other scenarios – because the systems that are changing and deciding things are not *separate* from humans, but are, rather, composite systems essentially involving humans.

So, in the global brain scenarios, one has more "human" empowerment than in some other cases – but the "humans" involved aren't legacy humans like us, but heavily networked humans that are largely characterized by the emergent dynamics and structures implicit in their interconnected activity!

12.10 Conclusion: Eight Ways to Bias AGI Toward Friendliness

It would be nice if we had a simple, crisp, comforting conclusion to this chapter on AGI ethics, but it's not the case. There is a certain irreducible uncertainty involved in creating advanced artificial minds. There is also a large irreducible uncertainty involved in the future of the human race in the case that we *don't* create advanced artificial minds: in accordance with the ancient Chinese curse, we live in interesting times!

What we can do, in this face of all this uncertainty, is to use our common sense to craft artificial minds that seem rationally and intuitively likely to be forces for good rather than otherwise – and revise our ideas frequently and openly based on what we learn as our research progresses. We have roughly outlined our views on AGI ethics, which have informed the CogPrime design in countless ways; but the current CogPrime design itself is just the initial condition for an AGI project. Assuming the project succeeds in creating an AGI preschooler, experimentation with this preschooler will surely teach us a great deal: both about AGI architecture in general, and about AGI ethics architecture in particular. We will then refine our cognitive and ethical theories and our AGI designs as we go about engineering, observing and teaching the next generation of systems.

All this is not a magic bullet for the creation of beneficial AGI systems, but we believe it's the right process to follow. The creation of AGI is part of a larger evolutionary process that human beings are taking part in, and the crafting of AGI ethics through engineering, interaction and instruction is also part of this process. There are no guarantees here – guarantees are rare in real life – but that doesn't mean that the situation is dire or hopeless, nor that (as some commentators have suggested [Joy00, McK03]) AGI research is too dangerous to pursue. It means we need to be mindful, intelligent, compassionate and cooperative as we proceed to carry out our parts in the next phase of the evolution of mind.

With this perspective in mind, we will conclude this chapter with a list of "Eight Ways to Bias Open-Source AGI Toward Friendliness", borrowed from a previous paper by Ben Goertzel and Joel Pitt of that name. These points summarize many of the points raised in the prior sections of this chapter, in a relatively crisp and practical manner:

1. **Engineer Multifaceted Ethical Capabilities**, corresponding to the multiple types of memory, including rational, empathic, imitative, etc.
2. **Foster Rich Ethical Interaction and Instruction**, with instructional methods according to the communication modes corresponding to all the types of memory: verbal, demonstrative, dramatic/depictive, indicative, goal-oriented.
3. **Engineer Stable, Hierarchy-Dominated Goal Systems** ... which is enabled nicely by CogPrime's goal framework and its integration with the rest of the CogPrime design
4. **Tightly Link AGI with the Global Brain**, so that it can absorb human ethical principles, both via natural interaction, and perhaps via practical implementations of current loosely-defined strategies like CEV, CAV and CBV
5. **Foster Deep, Consensus-Building Interactions Between People with Divergent Views**, so as to enable the interaction with the Global Brain to have the most clear and positive impact
6. **Create a Mutually Supportive Community of AGIs** which can then learn from each other and police against unfortunate developments (an approach which is meaningful if the AGIs are architected so as to militate against unexpected radical accelerations in intelligence)
7. **Encourage Measured Co-Advancement of AGI Software and AGI Ethics Theory**
8. **Develop Advanced AGI Sooner Not Later**

The last two of these points were not explicitly discussed in the body of the chapter, and so we will finalize the chapter by reviewing them here.

12.10.1 Encourage Measured Co-Advancement of AGI Software and AGI Ethics Theory

Everything involving AGI and Friendly AI (considered together or separately) currently involves significant uncertainty, and it seems likely that significant revision of current concepts will be valuable, as progress on the path toward powerful AGI proceeds. However, whether there is time for such revision to occur before AGI at the human level or above is created, depends on how fast is our progress toward AGI. What one wants is for progress to be slow enough that, at each stage of intelligence advance, concepts such as those discussed in this paper can be re-evaluated and re-analyzed in the light of the data gathered, and AGI designs and approaches can be revised accordingly as necessary.

However, due to the nature of modern technology development, it seems extremely unlikely that AGI development is going to be *artificially* slowed down in order to enable measured development of accompanying ethical tools, practices and understandings. For example, if one nation chose to enforce such a slowdown as a matter of policy (speaking about a future date at which substantial AGI progress has already been demonstrated, so that international AGI funding is dramatically increased from present levels), the odds seem very high that other nations would explicitly seek to accelerate their own progress on AGI, so as to reap the ensuing differential economic benefits (the example of stem cells arises again).

And this leads on to our next and final point regarding strategy for biasing AGI toward Friendliness....

12.10.2 Develop Advanced AGI Sooner Not Later

Somewhat ironically, it seems the best way to ensure that AGI development proceeds at a relatively measured pace is to *initiate serious AGI development sooner rather than later.* This is because the same AGI concepts will meet slower practical development today than 10 years from now, and slower 10 years from now than 20 years from now, etc. – due to the ongoing rapid advancement of various tools related to AGI development, such as computer hardware, programming languages, and computer science algorithms; and also the ongoing global advancement of education which makes it increasingly cost-effective to recruit suitably knowledgeable AI developers.

Currently the pace of AGI progress is sufficiently slow that practical work is in no danger of outpacing associated ethical theorizing. However, if we want to avoid the future occurrence of this sort of dangerous outpacing, our best practical choice is to make sure more substantial AGI development occurs in the phase *before* the development of tools that will make AGI development extraordinarily rapid. Of course, the authors are doing their best in this direction via their work on the CogPrime project!

Furthermore, this point bears connecting with the need, raised above, to foster the development of Global Brain technologies capable to "Foster Deep, Consensus-Building Interactions Between People with Divergent Views." If this sort of technology is to be maximally valuable, it should be created quickly enough that we can use it to help shape the goal system content of the first highly powerful AGIs. So, to simplify just a bit: We really want both deep-sharing GB technology and AGI technology to evolve relatively rapidly, compared to computing hardware and advanced CS algorithms (since the latter factors will be the main drivers behind the accelerating ease of AGI development). And this seems significantly challenging, since the latter receive dramatically more funding and focus at present.

If this perspective is accepted, then we in the AGI field certainly have our work cut out for us!

Section IV

Networks for Explicit and Implicit Knowledge Representation

Chapter 13

Local, Global and Glocal Knowledge Representation

Co-authored with Matthew Ikle, Joel Pitt and Rui Liu

13.1 Introduction

One of the most powerful metaphors we've found for understanding minds is to view them as **networks** – i.e. collections of interrelated, interconnected elements. The view of mind as network is implicit in the patternist philosophy, because every pattern can be viewed as a pattern *in* something, or a pattern of arrangement *of* something – thus a pattern is always viewable as a relation between two or more things. A collection of patterns is thus a pattern-network. Knowledge of all kinds may be given network representations; and cognitive processes may be represented as networks also; for instance via representing them as programs, which may be represented as trees or graphs in various standard ways. The emergent patterns arising in an intelligence as it develops may be viewed as a pattern network in themselves; and the relations between an embodied mind and its physical and social environment may be viewed in terms of ecological and social networks.

The chapters in this section are concerned with various aspects of networks, as related to intelligence in general and AGI in particular. Most of this material is not specific to CogPrime, and would be relevant to nearly any system aiming at human-level AGI. However, most of it has been developed in the course of work on CogPrime, and has direct relevance to understanding the intended operation of various aspects of a completed CogPrime system. We begin our excursion into networks, in this chapter, with an issue regarding networks and knowledge representation. One of the biggest decisions to make in designing an AGI system is how the system should represent knowledge. Naturally any advanced AGI system is going to synthesize a lot of its own knowledge representations for handling particular sorts of knowledge – but still, an AGI design typically makes *at least* some sort of commitment about the category of knowledge representation mechanisms toward which the AGI system will be biased. The two major supercategories of knowledge representation systems are *local* (also called *explicit*) and *global* (also called *implicit*) systems, with a hybrid category we refer to as *glocal* that combines both of these. In a local system, each piece of knowledge is stored using a small percentage of AGI system elements; in a global system, each piece of knowledge is stored using a particular pattern of arrangement, activation, etc. of a large percentage of AGI system elements; in a glocal system, the two approaches are used together.

In the first section here we discuss the symbolic, semantic-network aspects of knowledge representation in CogPrime

. Then we turn to distributed, neural-net-like knowledge representation, reviewing a host of general issues related to knowledge representation in attractor neural networks, turning finally to "glocal" knowledge representation mechanisms, in which ANNs combine localist and globalist representation, and explaining the relationship of the latter to CogPrime. The glocal aspect of CogPrime knowledge representation will become prominent in later chapters such as:

- in Chapter 23 of Part 2, where Economic Attention Networks (ECAN) are introduced and seen to have dynamics quite similar to those of the attractor neural nets considered here, but with a mathematics roughly modeling money flow in a specially constructed artificial economy rather than electrochemical dynamics of neurons.
- in Chapter 42 of Part 2, where "map formation" algorithms for creating localist knowledge from globalist knowledge are described

13.2 Localized Knowledge Representation using Weighted, Labeled Hypergraphs

There are many different mechanisms for representing knowledge in AI systems in an explicit, localized way, most of them descending from various variants of formal logic. Here we briefly describe how it is done in CogPrime, which on the surface is not that different from a number of prior approaches. (The particularities of CogPrime's explicit knowledge representation, however, are carefully tuned to match CogPrime's cognitive processes, which are more distinctive in nature than the corresponding representational mechanisms.)

13.2.1 Weighted, Labeled Hypergraphs

One useful way to think about CogPrime's explicit, localized knowledge representation is in terms of hypergraphs. A hypergraph is an abstract mathematical structure [Bol98], which consists of objects called Nodes and objects called Links which connect the Nodes. In computer science, a graph traditionally means a bunch of dots connected with lines (i.e. Nodes connected by Links). A hypergraph, on the other hand, can have Links that connect more than two Nodes.

In these pages we will often consider "generalized hypergraphs" that extend ordinary hypergraphs by containing two additional features:

- Links that point to Links instead of Nodes
- Nodes that, when you zoom in on them, contain embedded hypergraphs.

Properly, such "hypergraphs" should always be referred to as generalized hypergraphs, but this is cumbersome, so we will persist in calling them merely hypergraphs. In a hypergraph of this sort, Links and Nodes are not as distinct as they are within an ordinary mathematical graph (for instance, they can both have Links connecting them), and so it is useful to have a generic term encompassing both Links and Nodes; for this purpose, we use the term Atom.

A weighted, labeled hypergraph is a hypergraph whose Links and Nodes come along with labels, and with one or more numbers that are generically called weights. A label associated with a Link or Node may sometimes be interpreted as telling you what type of entity it is, or alternatively as telling you what sort of data is associated with a Node. On the other hand, an example of a weight that may be attached to an Link or Node is a number representing a probability, or a number representing how important the Node or Link is.

Obviously, hypergraphs may come along with various sorts of dynamics. Minimally, one may think about:

- Dynamics that modify the properties of Nodes or Links in a hypergraph (such as the labels or weights attached to them.)
- Dynamics that add new Nodes or Links to a hypergraph, or remove existing ones.

13.3 Atoms: Their Types and Weights

This section reviews a variety of CogPrime

Atom types and gives simple examples of each of them. The Atom types considered are drawn from those currently in use in the OpenCog system. This does not represent a complete list of Atom types referred to in the text of this book, nor a complete list of those used in OpenCog currently (though it does cover a substantial majority of those used in OpenCog currently, omitting only some with specialized importance or intended only for temporary use).

The partial nature of the list given here reflects a more general point: The specific collection of Atom types in an OpenCog system is bound to change as the system is developed and experiment with. CogPrime specifies a certain collection of representational approaches and cognitive algorithms for acting on them; any of these approaches and algorithms may be implemented with a variety of sets of Atom types. The specific set of Atom types in the OpenCog system currently does not necessarily have a profound and lasting significance – the list might look a bit different five years from time of writing, based on various detailed changes.

The treatment here is informal and intended to get across the general idea of what each Atom type does. A longer and more formal treatment of the Atom types is given in Part II, beginning in Chapter 20.

13.3.1 Some Basic Atom Types

We begin with ConceptNode – and note that a ConceptNode does not necessarily refer to a whole concept, but may refer to part of a concept – it is essentially a "basic semantic node" whose meaning comes from its links to other Atoms. It would be more accurately, but less tersely, named "concept or concept fragment or element node." A simple example would be a ConceptNode grouping nodes that are somehow related, e.g.

```
ConceptNode: C
InheritanceLink (ObjectNode: BW) C
InheritanceLink (ObjectNode: BP) C
InheritanceLink (ObjectNode: BN) C
ReferenceLink BW (PhraseNode "Ben's watch")
ReferenceLink BP (PhraseNode "Ben's passport")
ReferenceLink BN (PhraseNode "Ben's necklace")
```

indicates the simple and uninteresting ConceptNode grouping three objects owned by Ben (note that the above-given Atoms don't indicate the ownership relationship, they just link the three objects with textual descriptions). In this example, the ConceptNode links transparently to physical objects and English descriptions, but in general this won't be the case – most ConceptNodes will look to the human eye like groupings of links of various types, that link to other nodes consisting of groupings of links of various types, etc.

There are Atoms referring to basic, useful mathematical objects, e.g. NumberNodes like

```
NumberNode #4
NumberNode #3.44
```

The numerical value of a NumberNode is explicitly referenced within the Atom.

A core distinction is made between ordered links and unordered links; these are handled differently in the Atomspace software. A basic unordered link is the SetLink, which groups its arguments into a set. For instance, the ConceptNode C defined by

```
ConceptNode C
MemberLink A C
MemberLink B C
```

is equivalent to

```
SetLink A B
```

On the other hand, ListLinks are like SetLinks but ordered, and they play a fundamental role due to their relationship to predicates. Most predicates are assumed to take ordered arguments, so we may say e.g.

```
EvaluationLink
   PredicateNode eat
   ListLink
       ConceptNode cat
       ConceptNode mouse
```

to indicate that cats eat mice.

Note that by an expression like

```
ConceptNode cat
```

is meant

```
ConceptNode C
ReferenceLink W C
WordNode W #cat
```

since it's WordNodes rather than ConceptNodes that refer to words. (And note that the strength of the ReferenceLink would not be 1 in this case, because the word "cat" has multiple senses.) However, there is no harm nor formal incorrectness in the "ConceptNode cat" usage, since "cat" is just as valid a name for a ConceptNode as, say, "C."

We've already introduced above the MemberLink, which is a link joining a member to the set that contains it. Notable is that the truth value of a MemberLink is fuzzy rather than probabilistic, and that PLN is able to inter-operate fuzzy and probabilistic values.

SubsetLinks also exist, with the obvious meaning, e.g.

```
ConceptNode cat
ConceptNode animal
SubsetLink cat animal
```

Note that SubsetLink refers to a purely *extensional* subset relationship, and that InheritanceLInk should be used for the generic "intensional + extensional" analogue of this – more on this below. SubsetLink could more consistently (with other link types) be named ExtensionalInheritanceLink, but SubsetLink is used because it's shorter and more intuitive.

There are links representing Boolean operations AND, OR and NOT. For instance, we may say

```
ImplicationLink
    ANDLink
         ConceptNode young
         ConceptNode beautiful
     ConceptNode attractive
```

or, using links and VariableNodes instead of ConceptNodes,

```
AverageLink $X
   ImplicationLink
       ANDLink
            EvaluationLink young $X
            EvaluationLink beautiful $X
        EvaluationLink attractive $X
```

NOTLink is a unary link, so e.g. we might say

```
AverageLink $X
   ImplicationLink
       ANDLink
            EvaluationLink young $X
            EvaluationLink beautiful $X
            EvaluationLink
                 NOT
                 EvaluationLink poor $X
        EvaluationLink attractive $X
```

ContextLink allows explicit contextualization of knowledge, which is used in PLN, e.g.

```
ContextLink
   ConceptNode golf
   InheritanceLink
      ObjectNode BenGoertzel
      ConceptNode incompetent
```

says that Ben Goertzel is incompetent in the context of golf.

13.3.2 Variable Atoms

We have already introduced VariableNodes above; it's also possible to specify the type of a VariableNode via linking it to a VariableTypeNode via a TypedVariableLink, e.g.

```
VariableTypeLink
   VariableNode $X
   VariableTypeNode ConceptNode
```

which specifies that the variable $X should be filled with a ConceptNode.

Variables are handled via quantifiers; the default quantifier being the AverageLink, so that the default interpretation of

```
ImplicationLink
   InheritanceLink $X animal
   EvaluationLink
       PredicateNode: eat
       ListLink
           \$X
           ConceptNode: food
```

is

```
AverageLink $X
   ImplicationLink
      InheritanceLink $X animal
      EvaluationLink
          PredicateNode: eat
          ListLink
              \$X
              ConceptNode: food
```

The AverageLink invokes an estimation of the average TruthValue of the embedded expression (in this case an ImplicationLink) over all possible values of the variable $X. If there are type restrictions regarding the variable $X, these are taken into account in conducting the averaging. For AllLink and Exist s-Link may be used in the same places as AverageLink, with uncertain truth value semantics defined in PLN theory using third-order probabilities. There is also a ScholemLink used to indicate variable dependencies for existentially quantified variables, used in cases of multiply nested existential quantifiers.

EvaluationLink and MemberLink have overlapping semantics, allowing expression of the same conceptual/logical relationships in terms of predicates or sets, i.e.

```
EvaluationLink
   PredicateNode: eat
       ListLink
              $X
              ConceptNode: food
```

has the same semantics as

```
MemberLink
   ListLink
          $X
          ConceptNode: food
    ConceptNode: EatingEvents
```

The relation between the predicate "eat" and the concept "EatingEvents" is formally given by

```
ExtensionalEquivalenceLink
   ConceptNode: EatingEvents
   SatisfyingSetLink
         PredicateNode: eat
```

In other words, we say that "EatingEvents" is the SatisfyingSet of the predicate "eat": it is the set of entities that satisfy the predicate "eat". Note that the truth values of MemberLink and EvaluationLink are fuzzy rather than probabilistic.

13.3.3 Logical Links

There is a host of link types embodying logical relationships as defined in the PLN logic system, e.g.

- InheritanceLink
- SubsetLink (aka ExtensionalInheritanceLink)
- Intensional InheritanceLink

which embody different sorts of inheritance, e.g.

```
SubsetLink salmon fish
IntensionalInheritanceLink whale fish
InheritanceLink fish animal
```

and then

- SimilarityLink
- ExtensionalSimilarityLink
- IntensionalSimilarityLink

which are symmetrical versions, e.g.

```
SimilaritytLink shark barracuda
IntensionalSimilarityLink shark dolphin
ExtensionalSimiliarityLink American obese\_person
```

There are also higher-order versions of these links, both asymmetric

- ImplicationLink
- ExtensionalImplicationLink
- IntensionalImplicationLink

and symmetric

- EquivalenceLink
- ExtensionalEquivalenceLink
- IntensionalEquivalenceLink

These are used between predicates and links, e.g.

```
ImplicationLink
    EvaluationLink
         eat
        ListLink
             $X
             dirt
    EvaluationLink
         feel
         ListLInk
              $X
              sick
```

or

```
ImplicationLink
    EvaluationLink
         eat
        ListLink
             $X
             dirt
    InheritanceLink $X sick
```

or

```
ForAllLink $X, $Y, $Z
   ExtensionalEquivalenceLink
       EquivalenceLink
           $Z
          EvaluationLink
              +
              ListLink
                  $X
                  $Y
       EquivalenceLink
           $Z
          EvaluationLink
              +
              ListLink
                  $Y
                  $X
```

Note, the latter is given as an extensional equivalence because it's a pure mathematical equivalence. This is not the only case of pure extensional equivalence, but it's an important one.

13.3.4 Temporal Links

There are also temporal versions of these links, such as

- PredictiveImplicationLink
- PredictiveAttractionLink
- SequentialANDLink
- SimultaneousANDLink

which combine logical relation between the argument with temporal relation between their arguments. For instance, we might say

```
PredictiveImplicationLink
    PredicateNode: JumpOffCliff
    PredicateNode: Dead
```

or including arguments,

```
PredictiveImplicationLink
    EvaluationLink JumpOffCliff $X
    EvaluationLink Dead $X
```

The former version, without variable arguments given, shows the possibility of using higher-order logical links to join predicates without any explicit variables. Via using this format exclusively, one could avoid VariableAtoms entirely, using only higher-order functions in the manner of pure functional programming formalisms like combinatory logic. However, this purely functional style has not proved convenient, so the Atomspace in practice combines functional-style representation with variable-based representation.

Temporal links often come with specific temporal quantification, e.g.

```
PredictiveImplicationLink <5 seconds>
    EvaluationLink JumpOffCliff $X
    EvaluationLink Dead $X
```

indicating that the conclusion will generally follow the premise within 5 seconds. There is a system for managing fuzzy time intervals and their interrelationships, based on a fuzzy version of Allen Interval Algebra.

SequentialANDLink is similar to PredictiveImplicationLink but its truth value is calculated differently. The truth value of

```
SequentialANDLink <5 seconds>
    EvaluationLink JumpOffCliff $X
    EvaluationLink Dead $X
```

indicates the likelihood of the sequence of events occurring in that order, with gap lying within the specified time interval. The truth value of the PredictiveImplicationLink version indicates the likelihood of the second event, conditional on the occurrence of the first event (within the given time interval restriction).

There are also links representing basic temporal relationships, such as BeforeLink and AfterLink. These are used to refer to specific events, e.g. if X refers to the event of Ben waking up on July 15 2012, and Y refers to the event of Ben getting out of bed on July 15 2012, then one might have

```
AfterLink X Y
```

And there are TimeNodes (representing time-stamps such as temporal moments or intervals) and AtTimeLinks, so we may e.g. say

```
AtTimeLink
   X
   TimeNode: 8:24AM Eastern Standard Time, July 15 2012 AD
```

13.3.5 Associative Links

There are links representing associative, attentional relationships,

- HebbianLink
- AsymmetricHebbianLink
- InverseHebbianLink
- SymmetricInverseHebbianLink

These connote associations between their arguments, i.e. they connote that the entities represented by the two argument occurred in the same situation or context, for instance

```
HebbianLink happy smiling
AsymmetricHebbianLink dead rotten
InverseHebbianLink dead breathing
```

The asymmetric HebbianLink indicates that when the first argument is present in a situation, the second is also often present. The symmetric (default) version indicates that this relationship holds in both directions. The inverse versions indicate the negative relationship: e.g. when one argument is present in a situation, the other argument is often not present.

13.3.6 Procedure Nodes

There are nodes representing various sorts of procedures; these are kinds of ProcedureNode, e.g.

- SchemaNode, indicating any procedure
- GroundedSchemaNode, indicating any procedure associated in the system with a Combo program or C++ function allowing the procedure to be executed
- PredicateNode, indicating any predicate that associates a list of arguments with an output truth value
- GroundedPredicateNode, indicating a predicate associated in the system with a Combo program or C++ function allowing the predicate's truth value to be evaluated on a given specific list of arguments

ExecutionLinks and EvaluationLinks record the activity of SchemaNodes and PredicateNodes. We have seen many examples of EvaluationLinks in the above. Example ExecutionLinks would be:

```
ExecutionLink step\_forward
ExecutionLink step\_forward 5
ExecutionLink
    +
    ListLink
        NumberNode: 2
        NumberNode: 3
```

The first example indicates that the schema "step forward" has been executed. The second example indicates that it has been executed with an argument of "5" (meaning, perhaps, that 5 steps forward have been attempted). The last example indicates that the "+" schema has been executed on the argument list (2,3), presumably resulting in an output of 5.

The output of a schema execution may be indicated using an ExecutionOutputLink, e.g.

```
ExecutionOutputLink
    +
    ListLink
        NumberNode: 2
        NumberNode: 3
```

refers to the value "5" (as a NumberNode).

13.3.7 Links for Special External Data Types

Finally, there are also Atom types referring to specific types of data important to using OpenCog in specific contexts.

For instance, there are Atom types referring to general natural language data types, such as

- WordNode
- SentenceNode
- WordInstanceNode
- DocumentNode

plus more specific ones referring to relationships that are part of link-grammar parses of sentences

- FeatureNode
- FeatureLink
- LinkGrammarRelationshipNode
- LinkGrammarDisjunctNode

or RelEx semantic interpretations of sentences

- DefinedLinguisticConceptNode
- DefinedLinguisticRelationshipNode
- PrepositionalRelationshipNode

There are also Atom types corresponding to entities important for embodying OpenCog in a virtual world, e.g.

- ObjectNode
- AvatarNode
- HumanoidNode
- UnknownObjectNode
- AccessoryNode

13.3.8 Truth Values and Attention Values

CogPrime Atoms (Nodes and Links) are quantified with truth values that, in their simplest form, have two components, one representing probability (*strength*) and the other representing *weight of evidence*; and also with *attention values* that have two components, short-term and long-term importance, representing the estimated value of the Atom on immediate and long-term time-scales.

In practice many Atoms are labeled with CompositeTruthValues rather than elementary ones. A composite truth value contains many component truth values, representing truth values of the Atom in different contexts and according to different estimators.

It is important to note that the CogPrime declarative knowledge representation is neither a neural net nor a semantic net, though it does have some commonalities with each of these traditional representations. It is not a neural net because it has no activation values, and involves no attempts at low-level brain modeling. However, *attention values* are very loosely analogous to time-averages of neural net activations. On the other hand, it is not a semantic net because of the broad scope of the Atoms in the network: for example, Atoms may represent percepts, procedures, or parts of concepts. Most CogPrime Atoms have no corresponding English label. However, most CogPrime

Atoms do have probabilistic truth values, allowing logical semantics.

13.4 Knowledge Representation via Attractor Neural Networks

Now we turn to global, implicit knowledge representation – beginning with formal neural net models, briefly discussing the brain, and then turning back to CogPrime. Firstly, this section reviews some relevant material from the literature regarding the representation of knowledge using attractor neural nets. It is a mix of well-established fact with more speculative material.

13.4.1 The Hopfield neural net model

Hopfield networks [Hop82] are attractor neural networks often used as associative memories. A Hopfield network with N neurons can be trained to store a set of bipolar patterns P, where each pattern p has N bipolar (±1) values. A Hopfield net typically has symmetric weights with no self-connections. The weight of the connection between neurons i and j is denoted by w_{ij}.

In order to apply a Hopfield network to a given input pattern p, its activation state is set to the input pattern, and neurons are updated asynchronously, in random order, until the network converges to the closest fixed point. An often-used activation function for a neuron is:

$$y_i = sign(p_i \sum_{j \neq i} w_{ij} y_j)$$

Training a Hopfield network, therefore, involves finding a set of weights w_{ij} that stores the training patterns as attractors of its network dynamics, allowing future recall of these patterns from possibly noisy inputs.

Originally, Hopfield used a Hebbian rule to determine weights:

$$w_{ij} = \sum_{p=1}^{P} p_i p_j$$

Typically, Hopfield networks are fully connected. Experimental evidence, however, suggests that the majority of the connections can be removed without significantly impacting the network's capacity or dynamics. Our experimental work uses sparse Hopfield networks.

13.4.1.1 Palimpsest Hopfield nets with a modified learning rule

In [SV99] a new learning rule is presented, which both increases the Hopfield network capacity and turns it into a "palimpsest", i.e., a network that can continuously learn new patterns, while forgetting old ones in an orderly fashion.

Using this new training rule, weights are initially set to zero, and updated for each new pattern p to be learned according to:

$$h_{ij} = \sum_{k=1, k \neq i,j}^{N} w_{ik} p_k$$

$$\Delta w_{ij} = \frac{1}{n}(p_i p_j - h_{ij} p_j - h_{ji} p_i)$$

13.4.2 Knowledge Representation via Cell Assemblies

Hopfield nets and their ilk play a dual role: as computational algorithms, and as conceptual models of brain function. In CogPrime they are used as inspiration for slightly different, artificial economics based computational algorithms; but their hypothesized relevance to brain function is nevertheless of interest in a CogPrime context, as it gives some hints about the potential connection between low-level neural net mechanics and higher-level cognitive dynamics.

Hopfield nets lead naturally to a hypothesis about neural knowledge representation, which holds that a distinct mental concept is represented in the brain as either:

1. a set of "cell assemblies", where each assembly is a network of neurons that are interlinked in such a way as to fire in a (perhaps nonlinearly) synchronized manner
2. a distinct temporal activation pattern, which may occur in any one (or more) of a particular set of cell assemblies

For instance, this hypothesis is perfectly coherent if one interprets a "mental concept" as a SMEPH (defined in Chapter 14) ConceptNode, i.e. a fuzzy set of perceptual stimuli to which the organism systematically reacts in different ways. Also, although we will focus mainly on declarative knowledge here, we note that the same basic representational ideas can be applied to procedural and episodic knowledge: these may be hypothesized to correspond to temporal activation patterns as characterized above.

In the biology literature, perhaps the best-articulated modern theories championing the cell assembly view are those of Gunther Palm [Pal82, HAG07] and Susan Greenfield [SF05, CSG07]. Palm focuses on the dynamics of the formation and interaction assemblies of cortical columns. Greenfield argues that each concept has a core cell assembly, and that when the concept rises to the focus of attention, it recruits a number of other neurons beyond its core characteristic assembly into a "transient ensemble."[1]

It's worth noting that there may be multiple redundant assemblies representing the same concept – and potentially recruiting similar transient assemblies when highly activated. The importance of repeated, slightly varied copies of the same subnetwork has been emphasized by Edelman [Ede93] among other neural theorists.

---

[1] The larger an ensemble is, she suggests, the more vivid it is as a conscious experience; an hypothesis that accords well with the hypothesis made in [Goe06b] that a more informationally intense pattern corresponds to a more intensely conscious quale – but we don't need to digress extensively onto matters of consciousness for the present purposes.

13.5 Neural Foundations of Learning

Now we move from knowledge representation to learning – which is after all nothing but the adaptation of represented knowledge based on stimulus, reinforcement and spontaneous activity. While our focus in this chapter is on representation, it's not possible for us to make our points about glocal knowledge representation in neural net type systems without discussing some aspects of learning in these systems.

13.5.1 Hebbian Learning

The most common and plausible assumption about learning in the brain is that synaptic connections between neurons are adapted via some variant of Hebbian learning. The original Hebbian learning rule, proposed by Donald Hebb in his 1949 book [Heb49], was roughly

1. The weight of the synapse $x \rightarrow y$ increases if x and y fire at roughly the same time
2. The weight of the synapse $x \rightarrow y$ decreases if x fires at a certain time but y does not

Over the years since Hebb's original proposal, many neurobiologists have sought evidence that the brain actually uses such a method. One of the things they have found, so far, is a lot of evidence for the following learning rule [DC02, LS05]:

1. The weight of the synapse $x \rightarrow y$ increases if x fires shortly before y does
2. The weight of the synapse $x \rightarrow y$ decreases if x fires shortly after y does

The new thing here, not foreseen by Donald Hebb, is the "postsynaptic depression" involved in rule component 2.

Now, the simple rule stated above does not sum up all the research recently done on Hebbian-type learning mechanisms in the brain. The real biological story underlying these approximate rules is quite complex, involving many particulars to do with various neurotransmitters. Ill-understood details aside, however, there is an increasing body of evidence that not only does this sort of learning occur in the brain, but it leads to distributed experience-based neural modification: that is, one instance synaptic modification causes another instance of synaptic modification, which causes another, and so forth[2] [Bi01].

13.5.2 Virtual Synapses and Hebbian Learning Between Assemblies

Hebbian learning is conventionally formulated in terms of individual neurons, but, it can be extended naturally to assemblies via defining "virtual synapses" between assemblies.

Since assemblies are sets of neurons, one can view a synapse as linking two assemblies if it links two neurons, each of which is in one of the assemblies. One can then view two assemblies as being linked by a bundle of synapses. We can define the weight of the synaptic bundle from assembly A1 to assembly A2 as the number w so that *(the change*

---

[2] This has been observed in "model systems" consisting of neurons extracted from a brain and hooked together in a laboratory setting and monitored; measurement of such dynamics in vivo is obviously more difficult.

in the mean activation of A2 that occurs at time t+epsilon) is on average closest to $w \times$ *(the amount of energy flowing through the bundle from A1 to A2 at time t)*. So when A1 sends an amount x of energy along the synaptic bundle pointing from A1 to A2, then A2's mean activation is on average incremented/decremented by an amount $w \times x$.

In a similar way, one can define the weight of a bundle of synapses between a certain static or temporal activation-pattern P1 in assembly A1, and another static or temporal activation-pattern P2 in assembly A2. Namely, this may be defined as the number w so that *(the amount of energy flowing through the bundle from A1 to A2 at time t)*$\times w$ best approximates *(the probability that P2 is present in A2 at time t+epsilon)*, when averaged over all times t during which P1 is present in A1.

It is not hard to see that Hebbian learning on real synapses between neurons implies Hebbian learning on these virtual synapses between cell assemblies and activation-patterns.

These ideas may be developed further to build a connection between neural knowledge representation and probabilistic logical knowledge representation such as is used in CogPrime's Probabilistic Logic Networks formalism; this connection will be pursued at the end of Chapter 34, once more relevant background has been presented.

13.5.3 Neural Darwinism

A notion quite similar to Hebbian learning between assemblies has been pursued by Nobelist Gerald Edelman in his theory of neuronal group selection, or "Neural Darwinism." Edelman won a Nobel Prize for his work in immunology, which, like most modern immunology, was based on C. MacFarlane Burnet's theory of "clonal selection" [Bur62], which states that antibody types in the mammalian immune system evolve by a form of natural selection. From his point of view, it was only natural to transfer the evolutionary idea from one mammalian body system (the immune system) to another (the brain).

The starting point of Neural Darwinism is the observation that neuronal dynamics may be analyzed in terms of the behavior of neuronal groups. The strongest evidence in favor of this conjecture is physiological: many of the neurons of the neocortex are organized in clusters, each one containing say 10,000 to 50,000 neurons each. Once one has committed oneself to looking at such groups, the next step is to ask how these groups are organized, which leads to Edelman's concept of "maps."

A "map," in Edelman's terminology, is a connected set of groups with the property that when one of the inter-group connections in the map is active, others will often tend to be active as well. Maps are not fixed over the life of an organism. They may be formed and destroyed in a very simple way: the connection between two neuronal groups may be "strengthened" by increasing the weights of the neurons connecting the one group with the other, and "weakened" by decreasing the weights of the neurons connecting the two groups. If we replace "map" with "cell assembly" we arrive at a concept very similar to the one described in the previous subsection.

Edelman then makes the following hypothesis: *the large-scale dynamics of the brain is dominated by the natural selection of maps.* Those maps which are active when good results are obtained are strengthened, those maps which are active when bad results are obtained are weakened. And maps are continually mutated by the natural chaos of neural dynamics, thus providing new fodder for the selection process. By use of computer simulations, Edelman and his colleagues have shown that formal neural networks obeying this rule can carry out fairly complicated acts of perception. In general-evolution language, what is posited here is that organisms like humans contain chemical signals that signify organism-level success of various types, and that these signals serve as a "fitness function" correlating with evolutionary fitness of neuronal maps.

In *Neural Darwinism* and his other related books and papers, Edelman goes far beyond this crude sketch and presents neuronal group selection as a collection of precise biological hypotheses, and presents evidence in favor of a number of these hypotheses. However, we consider that the basic concept of neuronal group selection is largely independent of the biological particularities in terms of which Edelman has phrased it. We suspect that the mutation and selection of "transformations" or "maps" is a necessary component of the dynamics of any intelligent system.

As we will see later on (e.g. in Chapter 42 of Part 2, this business of maps is extremely important to CogPrime. CogPrime does not have simulated biological neurons and synapses, but it does have Nodes and Links that in some contexts play loosely similar roles. We sometimes think of CogPrime Nodes and Links as being very roughly analogous to Edelman's neuronal clusters, and emergent intercluster links. And we have maps among CogPrime Nodes and Links, just as Edelman has maps among his neuronal clusters. Maps are not the sole bearers of meaning in CogPrime, but they are significant ones.

There is a very natural connection between Edelman-style brain evolution and the ideas about cognitive evolution presented in Chapter 3. Edelman proposes a fairly clear mechanism via which patterns that survive a while in the brain are differentially likely to survive a long time: this is basic Hebbian learning, which in Edelman's picture plays a role between neuronal groups. And, less directly, Edelman's perspective also provides a mechanism by which intense patterns will be differentially selected in the brain: because on the level of neural maps, pattern intensity corresponds to the combination of compactness and functionality. Among a number of roughly equally useful maps serving the same function, the more compact one will be more likely to survive over time, because it is less likely to be disrupted by other brain processes (such as other neural maps seeking to absorb its component neuronal groups into themselves). Edelman's neuroscience remains speculative, since so much remains unknown about human neural structure and dynamics; but it does provide a tentative and plausible connection between evolutionary neurodynamics and the more abstract sort of evolution that patternist philosophy posits to occur in the realm of mind-patterns.

13.6 Glocal Memory

A *glocal* memory is one that transcends the global/local dichotomy and incorporates both aspects in a tightly interconnected way. Here we make the glocal memory concept more precise, and describe its incarnation in the context of attractor neural nets (which is similar to its incarnation in CogPrime, to be elaborated in later chapters). Though our main interest here is in glocality in CogPrime, we also suggest that glocality may be a critical property to consider when analyzing human, animal and AI memory more broadly.

The notion of glocal memory has implicitly occurred in a number of prior brain theories (without use of the neologism "glocal"), e.g. [Cal96] and [Goe01], but it has not previously been explicitly developed. However the concept has risen to the fore in our recent AI work and so we have chosen to flesh it out more fully in [HG08], [GPI$^+$10] and the present section.

Glocal memory overcomes the dichotomy between localized memory (in which each memory item is stored in a single location within an overall memory structure) and distributed memory (in which a memory item is stored as an aspect of a multi-component memory system, in such a way that the same set of multiple components stores a large number of memories). In a glocal memory system, most memory items are stored both locally and globally, with the property that eliciting either one of the two records of an item tends to also elicit the other one.

Glocal memory applies to multiple forms of memory; however we will focus largely on perceptual and declarative memory in our detailed analyses here, so as to conserve space and maintain simplicity of discussion.

The central idea of glocal memory is that (perceptual, declarative, episodic, procedural, etc.) items may be stored in memory in the form of paired structures that are called (key, map) pairs. Of course the idea of a "pair" is abstract, and such pairs may manifest themselves quite differently in different sorts of memory systems (e.g. brains versus non-neuromorphic AI systems). The key is a localized version of the item, and records some significant aspects of the items in a simple and crisp way. The map is a dispersed, distributed version of the item, which represents the item as a (to some extent, dynamically shifting) combination of fragments of other items. The map includes the key as a subset; activation of the key generally (but not necessarily always) causes activation of the map; and changes in the memory item will generally involve complexly coordinated changes on the key and map level both.

Memory is one area where animal brain architecture differs radically from the von Neumann architecture underlying nearly all contemporary general-purpose computers. Von Neumann computers separate memory from processing, whereas in the human brain there is no such distinction. In fact, it's arguable that in most cases the brain contains no memory apart from processing: human memories are generally constructed in the course of remembering [Ros88], which gives human memory a strong capability for "filling in gaps" of remembered experience and knowledge; and also causes problems with inaccurate remembering in many contexts [BF71, RM95] We believe the constructive aspect of memory is largely associated with its glocality.

The remainder of this section presents a fuller formalization of the glocal memory concept, which is then taken up further in three later chapters:

- Chapter **??** discusses the potential implementation of glocal memory in the human brain
- Chapter **??** discusses the implementation of glocal memory in attractor neural net systems
- Chapter 23 presents Glocal Economic Attention Networks (ECANs), rough analogues of glocal Hopfield nets that play a central role in CogPrime.

Our hypothesis of the potential **general** importance of glocality as a property of memory systems (beyond just the CogPrime architecture) – remains somewhat speculative. The presence of glocality in human and animal memory is strongly suggested but not firmly demonstrated by available neuroscience data; and the general value of glocality in the context of artificial brains and minds is also not yet demonstrated as the whole field of artificial brain and mind building remains in its infancy. However, the utility of glocal memory for CogPrime is not tied to this more general, speculative theme – glocality may be useful in CogPrime even if we're wrong that it plays a significant role in the brain and in intelligent systems more broadly.

13.6.1 A Semi-Formal Model of Glocal Memory

To explain the notion of glocal memory more precisely, we will introduce a simple semi-formal model of a system S that uses a memory to record information relevant to the actions it carries out. The overall concept of glocal memory should not be considered as restricted to this particular model. This model is not intended for maximal generality, but is intended to encompass a variety of current AI system designs and formal neurological models.

In this model, we will consider S's memory subsystem as a set of objects we'll call "tokens," embedded in some metric space. The metric in the space, which we will call the "basic distance" of the memory, generally will not be defined in terms of the semantics of the items stored in the memory; though it may come to shape these dynamics through the specific architecture and evolution of the memory. Note that these tokens are not intended as generally being mapped one-to-one onto meaningful items stored in the memory. The "tokens" are the raw materials that the memory arranges in various patterns in order to store items.

We assume that each token, at each point in time, may meaningfully be assigned a certain quantitative "activation level." Also, tokens may have other numerical or discrete quantities associated with them, depending on the particular memory architecture. Finally, tokens may relate other tokens, so that optionally a token may come equipped with an (ordered or unordered) list of other tokens.

To understand the meaning of the activation levels, one should think about S's memory subsystem as being coupled with an action-selection subsystem, that dynamically chooses the actions to be taken by the overall system in which the two subsystems are embedded. Each combination of actions, in each particular type of context, will generally be associated with the activation of certain tokens in memory.

Then, as analysts of the system S, we may associate each token T with an "activation vector" $v(T,t)$, whose value for each discrete time t consists of the activation of the token T at time t. So, the $50'th$ entry of the vector corresponds to the activation of the token at the $50'th$ time step.

"Items stored in memory" over a certain period of time, may then be defined as clusters in the set of activation vectors associated with memory during that period of time. Note that the system S itself may explicitly recognize and remember patterns regarding what items are stored in its memory – but, from an external analyst's perspective, the set of items in S's memory is not restricted to the ones that S has explicitly recognized as memory items.

The "localization" of a memory item may be defined as the degree to which the various tokens involved in the item are close to each other according to the metric in the memory metric-space. This degree may be formalized in various ways, but choosing a particular quantitative measure is not important here. A highly localized item may be called "local" and a not-very-localized item may be called "global."

We may define the "activation distance" of two tokens as the distance between their activation vectors. We may then say that a memory is "well aligned" to the extent that there is a correlation between the activation distance of tokens, and the basic distance of the memory metric-space.

Given the above set-up, the basic notion of glocal memory can be enounced fairly simply. A glocal memory is one:

- that is reasonably well-aligned (i.e. the correlation between activation and basic distance is significantly greater than random)

- in which most memory items come in pairs, consisting of one local item and one global item, so that activation of the local item (the "key") frequently leads in the near future to activation of the global item (the "map")

Obviously, in the scope of all possible memory structures constructible within the above formalism, glocal memories are going to be very rare and special. But, we suggest that they are important, because they are generally going to be the most effective way for intelligent systems to structure their memories.

Note also that many memories without glocal structure may be "well-aligned" in the above sense.

An example of a predominantly local memory structure, in which nearly all significant memory items are local according to the above definition, is the Cyc logical reasoning engine [LG90]. To cast the Cyc knowledge base in the present formal model, the tokens are logical predicates. Cyc does not have an in-built notion of activation, but one may conceive the activation of a logical formula in Cyc as the degree to which the formula is used in reasoning or query processing during a certain interval in time. And one may define a basic metric for Cyc by associating a predicate with its extension (the set of satisfying inputs), and defining the similarity of two predicates as the symmetric distance of their extensions. Cyc is reasonably well-aligned, but according to the dynamics of its querying and reasoning engines, it is basically a local memory structure without significant global memory structure.

On the other hand, an example of a predominantly global memory structure, in which nearly all significant memory items are global according to the above definition, is the Hopfield associative memory network [Ami89]. Here memories are stored in the pattern of weights associated with synapses within a network of formal neurons, and each memory in general involves a large number of the neurons in the network. To cast the Hopfield net in the present formal model, the tokens are neurons and synapses; the activations are neural net activations; the basic distance between two neurons A and B may be defined as the percentage of the time that stimulating one of the neurons leads to the other one firing; and to calculate a basic distance involving a synapse, one may associate the synapse with its source and target neurons. With these definitions, a Hopfield network is a well-aligned memory, and (by intentional construction) a markedly global one. Local memory items will be very rare in a Hopfield net.

While predominantly local and predominantly global memories may have great value for particular applications, our suggestion is that they also have inherent limitations. If so, this means that the most useful memories for general intelligence are going to be those that involve both local and global memory items in central roles. However, this is a more general and less risky claim than the assertion that glocal memory structure as defined above is important. Because, "glocal" as defined above doesn't just mean "neither predominantly global nor predominantly local." Rather, it refers to a specific pattern of coordination between local and global memory items – what we have called the "keys and maps" pattern.

13.6.2 Glocal Memory in the Brain

Science's understanding of human brain dynamics is still very primitive, one manifestation of which is the fact that we really don't understand how the brain represents knowledge, except in some very simple respects. So anything anyone says about knowledge representation in the brain, at this stage, has to be considered highly speculative. Existing neuroscience knowledge does imply constraints on how knowledge representation in the brain may work, but these are relatively loose constraints. These constraints do imply that, for instance, the brain is neither a relational database (in which information is stored in a wholly localized manner) nor a collection of "grandmother neurons" that respond individually to high-level percepts or concepts; nor a simple Hopfield type neural net (in which all memories are attractors globally distributed across the whole network). But they don't tell us nearly enough to, for instance, create a formal neural net model that can confidently be said to represent knowledge in the manner of the human brain.

As a first example of the current state of knowledge, we'll discuss here a series of papers regarding the neural representation of visual stimuli [QaGKKF05, QKKF08], which deal with the fascinating discovery of a subset of neurons in the medial temporal lobe (MTL) that are selectively activated by strikingly different pictures of given individuals, landmarks or objects, and in some cases even by letter strings. For instance, in their 2005 paper titled "Invariant visual representation by single neurons in the human brain", it is noted that

> in one case, a unit responded only to three completely different images of the ex-president Bill Clinton. Another unit (from a different patient) responded only to images of The Beatles, another one to cartoons from The Simpson's television series and another one to pictures of the basketball player Michael Jordan.

Their 2008 follow-up paper backed away from the more extreme interpretation in the title as well as the conclusion, with the title "Sparse but not 'Grandmother-cell' coding in the medial temporal lobe." As the authors emphasize there,

> Given the very sparse and abstract representation of visual information by these neurons, they could in principle be considered as 'grandmother cells'. However, we give several arguments that make such an extreme interpretation unlikely.
>
> ...
>
> MTL neurons are situated at the juncture of transformation of percepts into constructs that can be consciously recollected. These cells respond to percepts rather than to the detailed information falling on the retina. Thus, their activity reflects the full transformation that visual information undergoes through the ventral pathway. A crucial aspect of this transformation is the complementary development of both selectivity and invariance. The evidence presented here, obtained from recordings of single-neuron activity in humans, suggests that a subset of MTL neurons possesses a striking invariant representation for consciously perceived objects, responding to abstract concepts rather than more basic metric details. This representation is sparse, in the sense that responsive neurons fire only to very few stimuli (and are mostly silent except for their preferred stimuli), but it is far from a Grandmother-cell representation. The fact that the MTL represents conscious abstract information in such a sparse and invariant way is consistent with its prominent role in the consolidation of long-term semantic memories.

It's interesting to note how inadequate the [QKKF08] data really is for exploring the notion of glocal memory in the brain. Suppose it's the case that individual visual memories correspond to keys consisting of small neuronal subnetworks, and maps consisting of larger neuronal subnetworks. Then it would be not at all surprising if neurons in the "key" network corresponding to a visual concept like "Bill Clinton's face" would be found to respond differentially to the presentation of appropriate images. Yet, it would also be wrong to overinterpret such data as implying that the key network somehow comprises the "representation" of Bill Clinton's face in the individual's brain. In fact this key network would comprise only one aspect of said representation.

In the glocal memory hypothesis, a visual memory like "Bill Clinton's face" would be hypothesized to correspond to an attractor spanning a significant subnetwork of the individual's brain

– but this subnetwork still might occupy only a small fraction of the neurons in the brain (say, 1/100 or less), since there are very many neurons available. This attractor would constitute the map. But then, there would be a much smaller number of neurons serving as key to unlock this map: i.e. if a few of these key neurons were stimulated, then the overall attractor pattern in the map as a whole would unfold and come to play a significant role in the overall brain activity landscape. In prior publications [Goe97] the primary author explored this hypothesis in more detail in terms of the known architecture of the cortex and the mathematics of complex dynamical attractors.

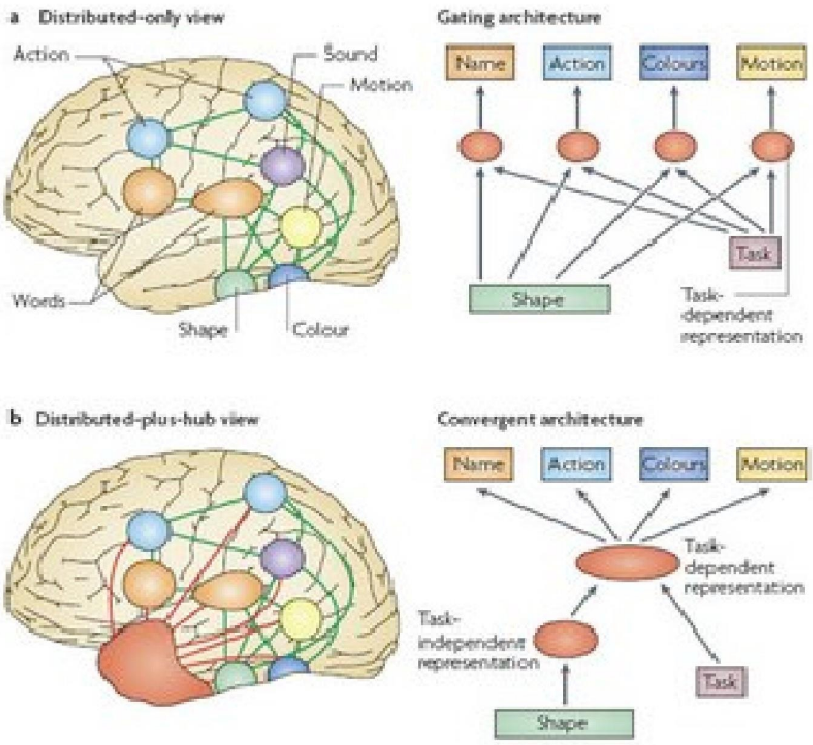
So, one possible interpretation of the [QKKF08] data is that the MTL neurons they're measuring are part of key networks that correspond to broader map networks recording percepts. The map networks might then extend more broadly throughout the brain, beyond the MTL and into other perceptual and cognitive areas of cortex. Furthermore, in this case, if some MTL key neurons were removed, the maps might well regenerate the missing keys (as would happen e.g. in the glocal Hopfield model to be discussed in the following section).

Related and interesting evidence for glocal memory in the brain comes from a recent study of semantic memory, illustrated in Figure **??** [PNR07]. Their research probed the architecture of semantic memory via comparing patients suffering from semantic dementia (SD) with patients suffering from three other neuropathologies, and found reasonably convincing evidence for what they call a "distributed-plus-hub" view of memory.

The SD patients they studied displayed highly distinctive symptomology; for instance, their vocabularies and knowledge of the properties of everyday objects were strongly impaired, whereas their memories of recent events and other cognitive capacities remain perfectly intact. These patients also showed highly distinctive patterns of brain damage: focal brain lesions in their anterior temporal lobes (ATL), unlike the other patients who had either less severe or more widely distributed damage in their ATLs. This led [PNR07] to conclude that the ATL (being adjacent to the amygdala and limbic systems that process reward and emotion; and the anterior parts of the medial temporal lobe memory system, which processes episodic memory) is a "hub" for amodal semantic memory, drawing general semantic information from episodic memories based on emotional salience.

So, in this view, the memory of something like a "banana" would contain a distributed aspect, spanning multiple brain systems, and also a localized aspect, centralized in the ATL. The distributed aspect would likely contain information on various particular aspects of bananas, including their sights, smells, and touches, the emotions they evoke, and the goals and motivations they relate to. The distributed and localized aspects would influence one another dynamically, but, the data [PNR07] gathered do not address dynamics and they don't venture hypotheses in this direction.

There is a relationship between the "distributed-plus-hub" view and [Dam00] better-known notion of a "convergence zone", defined roughly as a location where the brain binds features together. A convergence zone, in [Dam00] perspective, is not a "store" of information but an agent capable of decoding a signal (and of reconstructing information). He also uses the metaphor that convergence zones behave like indexes drawing information from other areas of the brain – but they are dynamic rather than static indices, containing the instructions needed to recognize and combine the features constituting the memory of something. The mechanism involved in the distributed-plus-hub model is similar to a convergence zone, but with the important difference that hubs are less local: [PNR07] semantic hub may be thought of a kind of "cluster of convergence zones" consisting of a network of convergence zones for various semantic memories.


a Distributed-only view
Action
Sound
Motion
Words
Shape
Colour
Gating architecture
Name
Action
Colours
Motion
Task
Shape
Task-dependent representation
b Distributed-plus-hub view
Convergent architecture
Name
Action
Colours
Motion
Task-dependent representation
Task-independent representation
Task
Shape

Nature Reviews | Neuroscience

Fig. 13.1: A Simplified Look at Feedback-Control in Uncertain Inference

What is missing in [PNR07] and [Dam00] perspective is a vision of distributed memories as attractors. The idea of localized memories serving as indices into distributed knowledge stores is important, but is only half the picture of glocal memory: the creative, constructive, dynamical-attractor aspect of the distributed representation is the other half. The closest thing to a clear depiction of this aspect of glocal memory that seems to exist in the neuroscience literature is a portion of William Calvin's theory of the "cerebral code" [Cal96]. Calvin proposes a set of quite specific mechanisms by which knowledge may be represented in the brain using complexly-structured strange attractors, and by which these strange attractors may be propagated throughout the brain. Figure 13.2 shows one aspect of his theory: how a distributed attractor may propagate from one part of the brain to another in pieces, with one portion of the attractor getting propagated first, and then seeding the formation in the destination brain region of a close approximation of the whole attractor.

Calvin's theory may be considered a genuinely glocal theory of memory. However, it also makes a large number of other specific commitments that are not part of the notion of glocality, such as his proposal of hexagonal meta-columns in the cortex, and his commitment to evolutionary learning as the primary driver of neural knowledge creation. We find these other

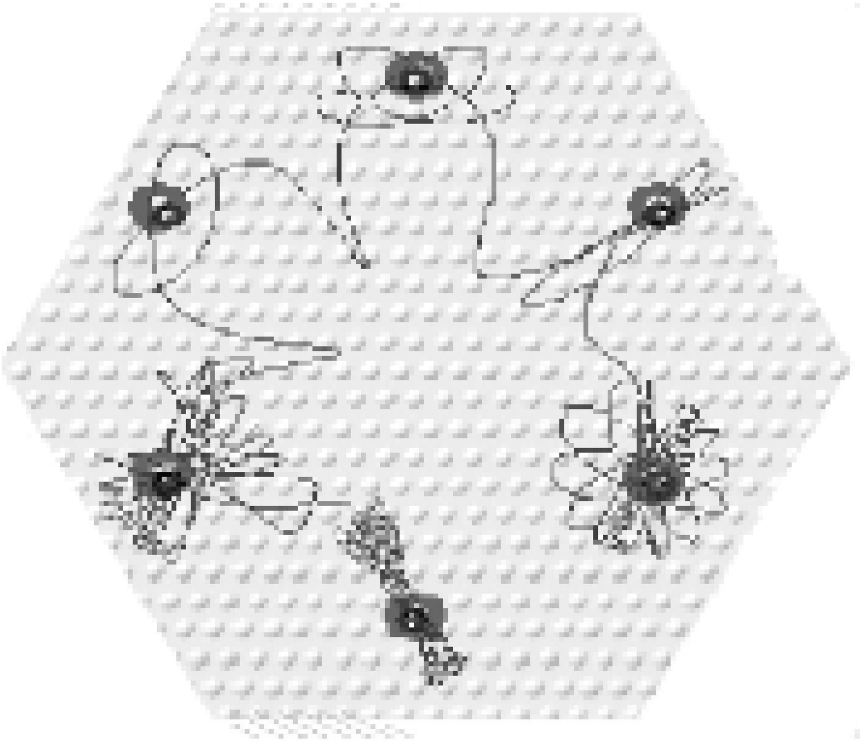


Fig. 13.2: Calvin's Model of Distributed Attractors in the Brain

hypotheses interesting and highly promising, yet feel it is also important to separate out the notion of glocal memory for separate consideration.

Regarding specifics, our suggestion is that Calvin's approach may overemphasize the distributed aspect of memory, not giving sufficient due to the relatively localized aspect as accounted for in the [QKKF08] results discussed above. In Calvin's glocal approach, global memories are attractors and local memories are parts of attractors. We suggest a possible alternative, in which global memories are attractors and local memories are particular neuronal subnetworks such as the specialized ones identified by [QKKF08]. However, this alternative does not seem contradictory to Calvin's overall conceptual approach, even though it is different from the particular proposals made in [Cal96].

The above paragraphs are far from a complete survey of the relevant neuroscience literature; there are literally dozens of studies one could survey pointing toward the glocality of various sorts of human memory. Yet experimental neuroscience tools are still relatively primitive, and every one of these studies could be interpreted in various other ways. In the next couple decades, as neuroscience tools improve in accuracy, our understanding of the role of glocality in human memory will doubtless improve tremendously.

13.6.3 Glocal Hopfield Networks

The ideas in the previous section suggest that, if one wishes to construct an AGI, it is worth seriously considering using a memory with some sort of glocal structure. One research direction that follows naturally from this notion is "glocal neural networks." In order to explore the nature of glocal neural networks in a relatively simple and tractable setting, we have formalized and implemented simple examples of "glocal Hopfield networks": palimpsest Hopfield nets with the addition of neurons representing localized memories. While these specific networks are not used in CogPrime, they are quite similar to the ECAN networks that are used in CogPrime and described in Chapter 23 of Part 2.

Essentially, we augment the standard Hopfield net architecture by adding a set of "key neurons." These are a small percentage of the neurons in the network, and are intended to be roughly equinumerous to the number of memories the network is supposed to store. When the Hopfield net converges to an attractor A, then new links are created between the neurons that are active in A, and one of the key neurons. Which key neuron is chosen? The one that, when it is stimulated, gives rise to an attractor pattern maximally similar to A.

The ultimate result of this is that, in addition to the distributed memory of attractors in the Hopfield net, one has a set of key neurons that in effect index the attractors. Each attractor corresponds to a single key neuron. In the glocal memory model, the key neurons are the keys and the Hopfield net attractors are the maps.

This algorithm has been tested in sparse Hopfield nets, using both standard Hopfield net learning rules and Storkey's modified palimpsest learning rule [SV99], which provides greater memory capacity in a continuous learning context. The use of key neurons turns out to slightly increase Hopfield net memory capacity, but this isn't the main point. The main point is that one now has a local representation of each global memory, so that if one wants to create a link between the memory and something else, it's extremely easy to do so – one just needs to link to the corresponding key neuron. Or, rather, one of the corresponding key neurons: depending on how many key neurons are allocated, one might end up with a number of key neurons corresponding to each memory, not just one.

In order to transform a palimpsest Hopfield net into a glocal Hopfield net, the following steps are taken:

1. Add a fixed number of "key neurons" to the network (removing other random neurons to keep the total number of neurons constant)
2. When the network reaches an attractor, create links from the elements in the attractor to one of the key neurons
3. The key neuron chosen for the previous step is the one that most closely matches the current attractor (which may be determined in several ways, to be discussed below)
4. To avoid the increase of the number of links in the network, when new links are created in Step 2, other key-neuron links are then deleted (several approaches may be taken here, but the simplest is to remove the key-neuron links with the lowest-absolute-value weights)

In the simple implementation of the above steps that we implemented, and described in [GPI+10], Step 3 is carried out simply by comparing the weights of a key neuron's links to the nodes in an attractor. A more sophisticated approach would be to select the key neuron with the highest activation during the transient interval immediately prior to convergence to the attractor.

The result of these modifications to the ordinary Hopfield net, is a Hopfield net that continually maintains a set of key neurons, each of which individually represents a certain attractor of the net.

Note that these key neurons – in spite of being "symbolic" in nature – are learned rather than preprogrammed, and are every bit as adaptive as the attractors they correspond to. Furthermore, if a key neuron is removed, the glocal Hopfield net algorithm will eventually learn it back, so the robustness properties of Hopfield nets are retained.

The results of experimenting with glocal Hopfield nets of this nature are summarized in [GPI+10]. We studied Hopfield nets with connectivity around .1, and in this context we found that glocality

- slightly increased memory capacity
- massively increased the rate of convergence to the attractor, i.e. the speed of recall

However, probably the most important consequence of glocality is a more qualitative one: it makes it far easier to link the Hopfield net into a larger system, as would occur if the Hopfield net were embedded in an integrative AGI architecture. Because a neuron external to the Hopfield net may now link to a memory in the Hopfield net by linking to the corresponding key neuron.

13.6.4 Neural-Symbolic Glocality in CogPrime

In CogPrime, we have explicitly sought to span the symbolic/emergentist pseudo-dichotomy, via creating an integrative knowledge representation that combines logic-based aspects with neural-net-like aspects. As reviewed in Chapter 6 above, these function not in the manner of multimodular systems, but rather via using (probabilistic) truth values and (attractor neural net like) attention values as weights on nodes and links of the same (hyper) graph. The nodes and links in this hypergraph are typed, like a standard semantic network approach for knowledge representation, so they're able to handle all sorts of knowledge, from the most concrete perception and actuation related knowledge to the most abstract relationships. But they're also weighted with values similar to neural net weights, and pass around quantities (importance values, discussed in Chapter 23 of Part 2) similar to neural net activations, allowing emergent attractor/assembly based knowledge representation similar to attractor neural nets.

The concept of glocality lies at the heart of this combination, in a way that spans the pseudo-dichotomy:

- Local knowledge is represented in abstract logical relationships stored in explicit logical form, and also in Hebbian-type associations between nodes and links.
- Global knowledge is represented in large-scale patterns of node and link weights, which lead to large-scale patterns of network activity, which often take the form of attractors qualitatively similar to Hopfield net attractors. These attractors are called *maps*.

The result of all this is that a concept like "cat" might be represented as a combination of:

- A small number of logical relationships and strong associations, that constitute the "key" subnetwork for the "cat" concept.
- A large network of weak associations, binding together various nodes and links of various types and various levels of abstraction, representing the "cat map".

The activation of the key will generally cause the activation of the map, and the activation of a significant percentage of the map will cause the activation of the rest of the map, including the key. Furthermore, if the key were for some reason forgotten, then after a significant amount of effort, the system would likely to be able to reconstitute it (perhaps with various small changes) from the information in the map. We conjecture that this particular kind of glocal memory will turn out to be very powerful for AGI, due to its ability to combine the strengths of formal logical inference with those of self-organizing attractor neural networks.

As a simple example, consider the representation of a "tower", in the context of an artificial agent that has built towers of blocks, and seen pictures of many other kinds of towers, and seen some tall building that it knows are somewhat like towers but perhaps not exactly towers. If this agent is reasonably conceptually advanced (say, at Piagetan the concrete operational level) then its mind will contain some declarative relationships partially characterizing the concept of "tower," as well as its sensory and episodic examples, and its procedural knowledge about how to build towers.

The key of the "tower" concept in the agent's mind may consist of internal images and episodes regarding the towers it knows best, the essential operations it knows are useful for building towers (piling blocks atop blocks atop blocks...), and the core declarative relations summarizing "towerness" – and the whole "tower" map then consists of a much larger number of images, episodes, procedures and declarative relationships connected to "tower" and other related entities. If any portion of the map is removed – even if the key is removed – then the rest of the map can be approximately reconstituted, after some work. Some cognitive operations are best done on the localized representation – e.g. logical reasoning. Other operations, such as attention allocation and guidance of inference control, are best done using the globalized map representation.

Chapter 14
Representing Implicit Knowledge via Hypergraphs

14.1 Introduction

Explicit knowledge is easy to write about and talk about; implicit knowledge is equally important, but tends to get less attention in discussions of AI and psychology, simply because we don't have as good a vocabulary for describing it, nor as good a collection of methods for measuring it. One way to deal with this problem is to describe implicit knowledge using language and methods typically reserved for explicit knowledge. This might seem intrinsically non-workable, but we argue that it actually makes a lot of sense. The same sort of networks that a system like CogPrime uses to represent knowledge explicitly, can also be used to represent the *emergent* knowledge that implicitly exists in an intelligent system's complex structures and dynamics.

We've noted that CogPrime uses an explicit representation of knowledge in terms of weighted labeled hypergraphs; and also uses other more neural net like mechanisms (e.g. the economic attention allocation network subsystem) to represent knowledge globally and implicitly. CogPrime combines these two sorts of representation according to the principle we have called *glocality*. In this chapter we pursue glocality a bit further – describing a means by which even implicitly represented knowledge can be modeled using weighted labeled hypergraphs similar to the ones used explicitly in CogPrime. This is conceptually important, in terms of making clear the fundamental similarities and differences between implicit and explicit knowledge representation; and it is also pragmatically meaningful due to its relevance to the CogPrime methods described in Chapter 42 of Part 2 that transform implicit into explicit knowledge.

To avoid confusion with CogPrime's explicit knowledge representation, we will refer to the hypergraphs in this chapter as composed of Vertices and Edges rather than Nodes and Links. In prior publications we have referred to "derived" or "emergent" hypergraphs of the sort described here using the acronym **SMEPH**, which stands for Self-Modifying, Evolving Probabilistic Hypergraphs.

14.2 Key Vertex and Edge Types

We begin by introducing a particular collection of Vertex and Edge types, to be used in modeling the internal structures of intelligent systems.

The key SMEPH Vertex types are

- ConceptVertex, representing a set, for instance, an idea or a set of percepts
- SchemaVertex, representing a procedure for doing something (perhaps something in the physical world, or perhaps an abstract mental action).

The key SMEPH Edge types, using language drawn from Probabilistic Logic Networks (PLN) and elaborated in Chapter 34 below, are as follows:

- ExtensionalInheritanceEdge (ExtInhEdge for short: an edge which, linking one Vertex or Edge to another, indicates that the former is a special case of the latter)
- ExtensionalSimilarityEdge (ExtSim: which indicates that one Vertex or Edge is similar to another)
- ExecutionEdge (a ternary edge, which joins S,B,C when S is a SchemaVertex and the result from applying S to B is C).

So, in a SMEPH system, one is often looking at hypergraphs whose Vertices represent ideas or procedures, and whose Edges represent relationships of specialization, similarity or transformation among ideas and/or procedures.

The semantics of the SMEPH edge types is given by PLN, but is simple and commonsensical. ExtInh and ExtSim Edges come with probabilistic weights indicating the extent of the relationship they denote (e.g. the ExtSimEdge joining the cat ConceptVertex to the dog ConceptVertex gets a higher probability weight than the one joining the cat ConceptVertex to the washing-machine ConceptVertex). The mathematics of transformations involving these probabilistic weights becomes quite involved - particularly when one introduces SchemaVertices corresponding to abstract mathematical operations, a step that enables SMEPH hypergraphs to have the complete mathematical power of standard logical formalisms like predicate calculus, but with the added advantage of a natural representation of uncertainty in terms of probabilities, as well as a natural representation of networks and webs of complex knowledge.

14.3 Derived Hypergraphs

We now describe how SMEPH hypergraphs may be used to model and describe intelligent systems. One can (in principle) draw a SMEPH hypergraph corresponding to any individual intelligent system, with Vertices and Edges for the concepts and processes in that system's mind. This is called the derived hypergraph of that system.

14.3.1 SMEPH Vertices

A ConceptVertex in the derived hypergraph of a system corresponds to a structural pattern that persists over time in that system; whereas a SchemaVertex corresponds to a multi-time-point dynamical pattern that recurs in that system's dynamics. If one accepts the patternist definition of a mind as the set of patterns in an intelligent system, then it follows that the derived hypergraph of an intelligent system captures a significant fraction of the mind of that system.

To phrase it a little differently, we may say that a ConceptVertex, in SMEPH, refers to the habitual pattern of activity observed in a system when some condition is met (this condition corresponding to the presence of a certain pattern). The condition may refer to something in the world external to the system, or to something internal. For instance, the condition may be observing a cat. In this case, the corresponding Concept vertex in the mind of Ben Goertzel is the pattern of activity observed in Ben Goertzel's brain when his eyes are open and he's looking in the direction of a cat. The notion of pattern of activity can be made rigorous using mathematical pattern theory, as is described in The Hidden Pattern [Goe06a].

Note that logical predicates, on the SMEPH level, appear as particular kinds of Concepts, where the condition involves a predicate and an argument. For instance, suppose one wants to know what happens inside Ben's mind when he eats cheese. Then there is a Concept corresponding to the condition of cheese-eating activity. But there may also be a Concept corresponding to eating activity in general. If the Concept denoting the activity of eating X is generally easily computable from the Concepts for X and eating individually, then the eating Concept is effectively acting as a predicate.

A SMEPH SchemaVertex, on the other hand, is like a Concept that's defined in a time-dependent way. One type of Schema refers to a habitual dynamical pattern of activity occurring before and/or during some condition is met. For instance, the condition might be saying the word Hello. In that case the corresponding SchemaVertex in the mind of Ben Goertzel is the pattern of activity that generally occurs before he says Hello.

Another type of Schema refers to a habitual dynamical pattern of activity occurring after some condition X is met. For instance, in the case of the Schema for adding two numbers, the precondition X consists of the two numbers and the concept of addition. The Schema is then what happens when the mind thinks of adding and thinks of two numbers.

Finally, there are Schema that refer to habitual dynamical activity patterns occurring after some condition X is met and before some condition Y is met. In this case the Schema is viewed as transforming X into Y. For instance, if X is the condition of meeting someone who is not a friend, and Y is the condition of being friends with that person, then the habitually intervening activities constitute the Schema for making friends.

14.3.2 SMEPH Edges

SMEPH edge types fall into two categories: functional and logical. Functional edges connect Schema vertices to their input and outputs; logical edges refer mainly to conditional probabilities, and in general are to be interpreted according to the semantics of Probabilistic Logic Networks.

Let us begin with logical edges. The simplest case is the Subset edge, which denotes a straightforward, extensional conditional probability. For instance, it may happen that whenever the Concept for cat is present in a system, the Concept for animal is as well. Then we would say

```
Subset cat animal
```

(Here we assume a notation where "R A B" denotes an Edge of type R between Vertices A and B.)

On the other hand, it may be that 50% of the time that cat is present in the system, cute is present as well: then we would say

```
Subset cat cute <.5>
```

where the <.5> denotes the probability, which is a component of the Truth Value associated with the edge.

Next, the most basic functional edge is the Execution edge, which is ternary and denotes a relation between a Schema, its input and its output, e.g.

```
Execution father_of Ben_Goertzel Ted_Goertzel
```

for a schema father_of that outputs the father of its argument.

The ExecutionOutput (ExOut) edge denotes the output of a Schema in an implicit way, e.g.

```
ExOut say_hello
```

refers to a particular act of saying hello, whereas

```
ExOut add_numbers {3, 4)
```

refers to the Concept corresponding to 7. Note that this latter example involves a set of three entities: sets are also part of the basic SMEPH knowledge representation. A set may be thought of as a hypergraph edge that points to all its members.

In this manner we may define a set of edges and vertices modeling the habitual activity patterns of a system when in different situations. This is called the derived hypergraph of the system. Note that this hypergraph can in principle be constructed no matter what happens inside the system: whether it's a human brain, a formal neural network, Cyc, OCP, a quantum computer, etc. Of course, constructing the hypergraph in practice is quite a different story: for instance, we currently have no accurate way of measuring the habitual activity patterns inside the human brain. fMRI and PET and other neuroimaging technologies give only a crude view, though they are continually improving.

Pattern theory enters more deeply here when one thoroughly fleshes out the Inheritance concept. Philosophers of logic have extensively debated the relationship between extensional inheritance (inheritance between sets based on their members) and intensional inheritance (inheritance between entity-types based on their properties). A variety of formal mechanisms have been proposed to capture this conceptual distinction; see (Wang, 2006, 1995 TODO make ref) for a review along with a novel approach utilizing uncertain term logic. Pattern theory provides a novel approach to defining intension: one may associate with each ConceptVertex in a system's derived hypergraph the set of patterns associated with the structural pattern underlying that ConceptVertex. Then, one can define the strength of the IntensionalInheritanceEdge between two ConceptVertices A and B as the percentage of A's pattern-set that is also contained in B's pattern-set. According to this approach, for instance, one could have

```
IntInhEdge whale fish <0.6>

ExtInhEdge whale fish <0.0>
```

since the fish and whale sets have common properties but no common members.

14.4 Implications of Patternist Philosophy for Derived Hypergraphs of Intelligent Systems

Patternist philosophy rears its head here and makes some definite hypotheses about the structure of derived hypergraphs. It suggests that derived hypergraphs should have a dual network structure, and that in highly intelligent systems they should have subgraphs that constitute models of the whole hypergraph (these are self systems). SMEPH does not add anything to the patternist view on a philosophical level, but it gives a concrete instantiation to some of the general ideas of patternism. In this section we'll articulate some "SMEPH principles", constituting important ideas from patternist philosophy as they manifest themselves in the SMEPH context.

The logical edges in a SMEPH hypergraph are weighted with probabilities, as in the simple example given above. The functional edges may be probabilistically weighted as well, since some Schema may give certain results only some of the time. These probabilities are critical in terms of SMEPH's model of system dynamics; they underly one of our SMEPH principles,

Principle of Implicit Probabilistic Inference: In an intelligent system, the temporal evolution of the probabilities on the edges in the system's derived hypergraph should approximately obey the rules of probability theory.

The basic idea is that, even if a system - through its underlying dynamics - has no explicit connection to probability theory, it still must behave roughly as if it does, if it is going to be intelligent. The roughly part is important here; it's well known that humans are not terribly accurate in explicitly carrying out formal probabilistic inferences. And yet, in practical contexts where they have experience, humans can make quite accurate judgments; which is all that's required by the above principle, since it's the contexts where experience has occurred that will make up a system's derived hypergraph.

Our next SMEPH principle is evolutionary, and states

Principle of Implicit Evolution: In an intelligent system, new Schema and Concepts will continually be created, and the Schema and Concepts that are more useful for achieving system goals (as demonstrated via probabilistic implication of goal achievement) will tend to survive longer.

Note that this principle can be fulfilled in many different ways. The important thing is that system goals are allowed to serve as a selective force.

Another SMEPH dynamical principle pertains to a shorter time-scale than evolution, and states

Principle of Attention Allocation: In an intelligent system, Schema and Concepts that are more useful for attaining short-term goals will tend to consume more of the system's energy. (The balance of attention oriented toward goals pertaining to different time scales will vary from system to system.)

Next, there is the

Principle of Autopoesis: In an intelligent system, if one removes some part of the system and then allows the system's natural dynamics to keep going, a decent approximation to that removed part will often be spontaneously reconstituted.

And there is the

Cognitive Equation Principle: In an intelligent system, many abstract patterns that are present in the system at a certain time as patterns among other Schema and Concepts, will at a near-future time be present in the system as patterns among elementary system components.

The Cognitive Equation Principle, briefly discussed in Chapter 3, basically means that Concepts and Schema emergent in the system are recognized by the system and then embodied as elementary items in the system so that patterns among them in their emergent form become, with the passage of time, patterns among them in their directly-system-embodied form. This is a natural consequence of the way intelligent systems continually recognize patterns in themselves.

Note that derived hypergraphs may be constructed corresponding to any complex system which demonstrates a variety of internal dynamical patterns depending on its situation. However, if a system is not intelligent, then according to the patternist philosophy evolution of its derived hypergraph can't necessarily be expected to follow the above principles.

14.4.1 SMEPH Principles in CogPrime

We now more explicitly elaborate the application of these ideas in the CogPrime context. As noted above, in addition to explicit knowledge representation in terms of Nodes and Links, CogPrime also incorporates implicit knowledge representation in the form of what are called Maps: collections of Nodes and Links that tend to be utilized together within cognitive processes.

These Maps constitute a CogPrime system's derived hypergraph, which will not be identical to the hypergraph it uses for explicit knowledge representation. However, an interesting feedback loop arises here, in that the intelligence's self-study will generally lead it to recognize large portions of its derived hypergraph as patterns in itself, and then embody these patterns within its concretely implemented knowledge hypergraph. This relates to the Cognitive Equation Principle defined above 3, in which an intelligent system continually recognizes patterns in itself and embodies these patterns in its own basic structure (so that new patterns may more easily emerge from them).

Often it happens that a particular CogPrime node will serve as the center of a map, so that e.g. the Concept Link denoting cat will consist of a number of nodes and links roughly centered around a ConceptNode that is linked to the WordNode cat. But this is not guaranteed and some CogPrime maps are more diffuse than this with no particular center.

Somewhat similarly, the key SMEPH dynamics are represented explicitly in CogPrime: probabilistic reasoning is carried out via explicit application of PLN on the CogPrime hypergraph, evolutionary learning is carried out via application of the MOSES optimization algorithm, and attention allocation is carried out via a combination of inference and evolutionary pattern mining. But the SMEPH dynamics also occur implicitly in CogPrime: emergent maps are reasoned on probabilistically as an indirect consequence of node-and-link level PLN activity; maps evolve as a consequence of the coordinated whole of CogPrime dynamics; and attention shifts between maps according to complex emergent dynamics.

To see the need for maps, consider that even a Node that has a particular meaning attached to it - like the *Iraq* Node, say - doesn't contain much of the meaning of *Iraq* in it. The meaning of *Iraq* lies in the Links attached to this Node, and the Links attached to their Nodes - and the other Nodes and Links not explicitly represented in the system, which will be created by

CogPrime's cognitive algorithms based on the explicitly existent Nodes and Links related to the *Iraq* Node.

This halo of Atoms related to the *Iraq* node is called the *Iraq* map. In general, some maps will center around a particular Atom, like this *Iraq* map, others may not have any particular identifiable center. CogPrime's cognitive processes act directly on the level of Nodes and Links, but they must be analyzed in terms of their impact on maps as well. In SMEPH terms, CogPrime maps may be said to correspond to SMEPH ConceptNodes, and for instance bundles of Links between the Nodes belonging to a map may correspond to a SMEPH Link between two ConceptNodes.

Chapter 15
Emergent Networks of Intelligence

15.1 Introduction

When one is involved with engineering an AGI system, one thinks a lot about the aspects of the system one is explicitly building – what are the parts, how they fit together, how to test they're properly working, and so forth. And yet, these explicitly engineered aspects are only a fraction of what's important in an AGI system. At least as critical are the *emergent* aspects – the patterns that emerge once the system is up and running, interacting with the world and other agents, growing and developing and learning and self-modifying. SMEPH is one toolkit for describing some of these emergent patterns, but it's only a start.

In line with these general observations, most of this book will focus on the structures and processes that we have built, or intend to build, into the CogPrime system. But in a sense, these structures and processes are not the crux of CogPrime's intended intelligence. The purpose of these pre-programmed structures and processes is to give rise to *emergent* structures and processes, in the course of CogPrime's interaction with the world and the other minds within it. We will return to this theme of emergence at several points in later chapters, e.g. in the discussion of map formation in Chapter 42 of Part 2.

Given the important of emergent structures – and specifically emergent *network* structures – for intelligence, it's fortunate the scientific community has already generated a lot of knowledge about complex networks: both networks of physical or software elements, and networks of organization emergent from complex systems. As most of this knowledge has originated in fields other than AGI, or in pure mathematics, it tends to require some reinterpretation or tweaking to achieve maximal applicability in the AGI context; but we believe this effort will become increasingly worthwhile as the AGI field progresses, because network theory is likely to be very useful for describing the contents and interactions of AGI systems as they develop increasing intelligence.

In this brief chapter we specifically focus on the emergence of certain large-scale *network structures* in a CogPrime knowledge store, presenting heuristic arguments as to why these structures can be expected to arise. We also comment on the way in which these emergent structures are expected to guide cognitive processes, and give rise to emergent cognitive processes. The following chapter expands on this theme in a particular direction, exploring the possible emergence of structures characterizing inter-cognitive reflection.

15.2 Small World Networks

One simple but potentially useful observation about CogPrime Atomspaces is that they are generally going to be *small world networks* [Buc03], rather than random graphs. A small world network is a graph in which the connectivities of the various nodes display a power law behavior – so that, loosely speaking, there are a few nodes with very many links, then more nodes with a modest number of links ... and finally, a huge number of nodes with very few links. This kind of network occurs in many natural and human systems, including citations among papers, financial arrangements among banks, links between Web pages and the spread of diseases among people or animals. In a weighted network like an Atomspace, "small-world-ness" must be defined in a manner taking the weights into account, and there are several obvious ways to do this. Figure 15.1 depicts a small but prototypical small-worlds network, with a few "hub" nodes possessing far more neighbors than the others, and then some secondary hubs, etc.

An excellent reference on network theory in general, including but not limited to small world networks, is Peter Csermely's *Weak Links* [Cse06]. Many of the ideas in that work have apparent OpenCog applications, which are not elaborated here.

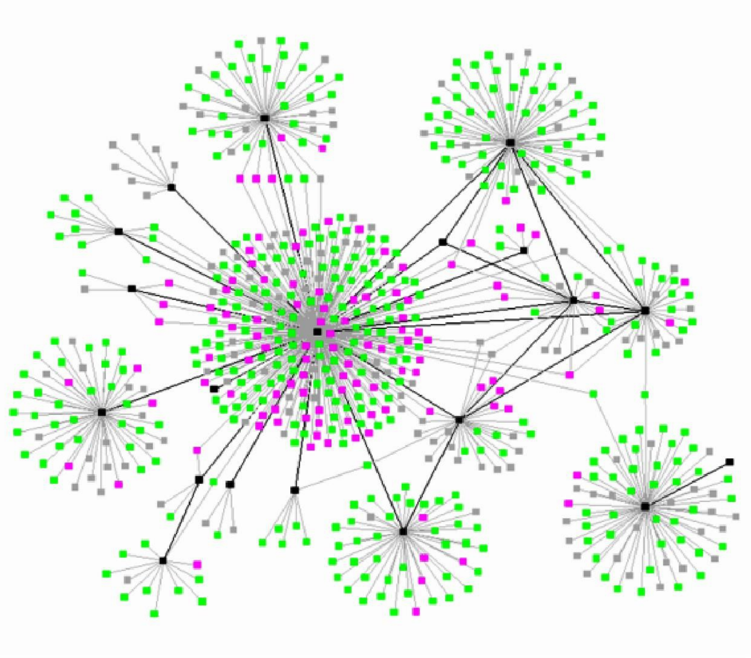


Fig. 15.1: A typical, though small-sized, small-worlds network.

One process via which small world networks commonly form is "preferential attachment" [Bar02]. This occurs in essence when "the rich get richer" – i.e. when nodes in the network grow new links, in a manner that causes them to preferentially grow links to nodes that already have more links. It is not hard to see that CogPrime's ECAN dynamics will naturally lead to preferential attachment, because Atoms with more links will tend to get more STI, and thus will tend to get selected by more cognitive processes, which will cause them to grow more links. For this reason, in most circumstances, a CogPrime system in which most link-building cognitive processes rely heavily on ECAN to guide their activities will tend to contain a small-world-network Atomspace. This is not rigorously guaranteed to be the case for any possible combination of environment and goals, but it is commonsensically likely to nearly always be the case.

One consequence of the small worlds structure of the Atomspace is that, in exploring other properties of the Atom network, it is particularly important to look at the hub nodes. For instance, if one is studying whether hierarchical and heterarchical subnetworks of the Atomspace exist, and whether they are well-aligned with each other, it is important to look at hierarchical and heterarchical connections between hub nodes in particular (and secondary hubs, etc.). A pattern of hierarchical or dual network connection that only held up among the more sparsely connected nodes in a small-world network would be a strange thing, and perhaps not that cognitively useful.

15.3 Dual Network Structure

One of the key theoretical notions in patternist philosophy is that complex cognitive systems evolve internal *dual network* structures, comprising superposed, harmonized hierarchical and heterarchical networks. Now we explore some of the specific CogPrime structures and dynamics militating in favor of the emergence of dual networks.

15.3.1 Hierarchical Networks

The hierarchical nature of human linguistic concepts is well known, and is illustrated in Figure 15.2 for the commonsense knowledge domain (using a graph drawn from WordNet, a huge concept hierarchy covering 50K+ English-language concepts), and in Figure 15.4 for a specialized knowledge subdomain, genetics. Due to this fact, a certain amount of hierarchy can be expected to emerge in the Atomspace of any linguistically savvy CogPrime, simply due to its modeling of the linguistic concepts that it hears and reads.

Hierarchy also exists in the natural world apart from language, which is the reason that many sensorimotor-knowledge-focused AGI systems (e.g. DeSTIN and HTM, mentioned in Chapter 4 above) feature hierarchical structures. In these cases the hierarchies are normally spatiotemporal in nature - with lower layers containing elements responding to more localized aspects of the perceptual field, and smaller, more localized groups of actuators. This kind of hierarchy certainly *could* emerge in an AGI system, but in CogPrime we have opted for a different route. If a CogPrime system is hybridized with a hierarchical sensorimotor network like one of those mentioned above, then the Atoms linked to the nodes in the hierarchical sensorimotor network will naturally possess hierarchical conceptual relationships, and will thus naturally grow hierarchical links between them (e.g. InheritanceLinks and IntensionalInheritanceLinks via PLN, AsymmetricHebbianLinks via ECAN).

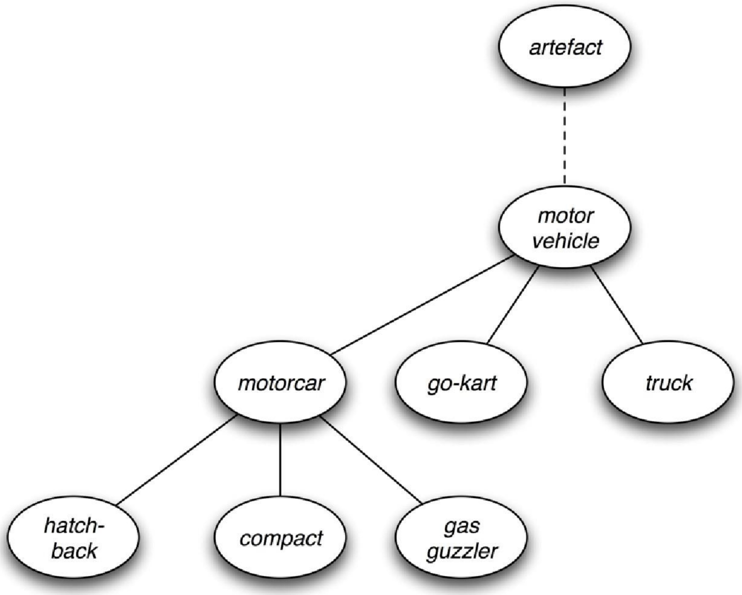


Fig. 15.2: A typical, though small, subnetwork of WordNet's hierarchical network.

Once elements of hierarchical structure exist via the hierarchical structure of language and physical reality, then a richer and broader hierarchy can be expected to accumulate on top of it, because importance spreading and inference control will implicitly and automatically be guided by the existing hierarchy. That is, in the language of *Chaotic Logic* [Goe94] and patternist theory, hierarchical structure is an "autopoietic attractor" – once it's there it will tend to enrich itself and maintain itself. AsymmetricHebbianLinks arranged in a hierarchy will tend to cause importance to spread up or down the hierarchy, which will lead other cognitive processes to look for patterns between Atoms and their hierarchical parents or children, thus potentially building more hierarchical links. Chains of InheritanceLinks pointing up and down the hierarchy will lead PLN to search for more hierarchical links – e.g. most simply, $A \to B \to C$ where C is above B is above A in the hierarchy, will naturally lead inference to check the viability of $A \to C$ by deduction. There is also the possibility to introduce a special DefaultInheritanceLink, as discussed in Chapter 34 of Part 2, but this isn't actually necessary to obtain the inferential maintenance of a robust hierarchical network.

15.3.2 *Associative, Heterarchical Networks*

Heterarchy is in essence a simpler structure than hierarchy: it simply refers to a network in which nodes are linked to other nodes with which they share important relationships. That is, there should be a tendency that if two nodes are often important in the same contexts or for

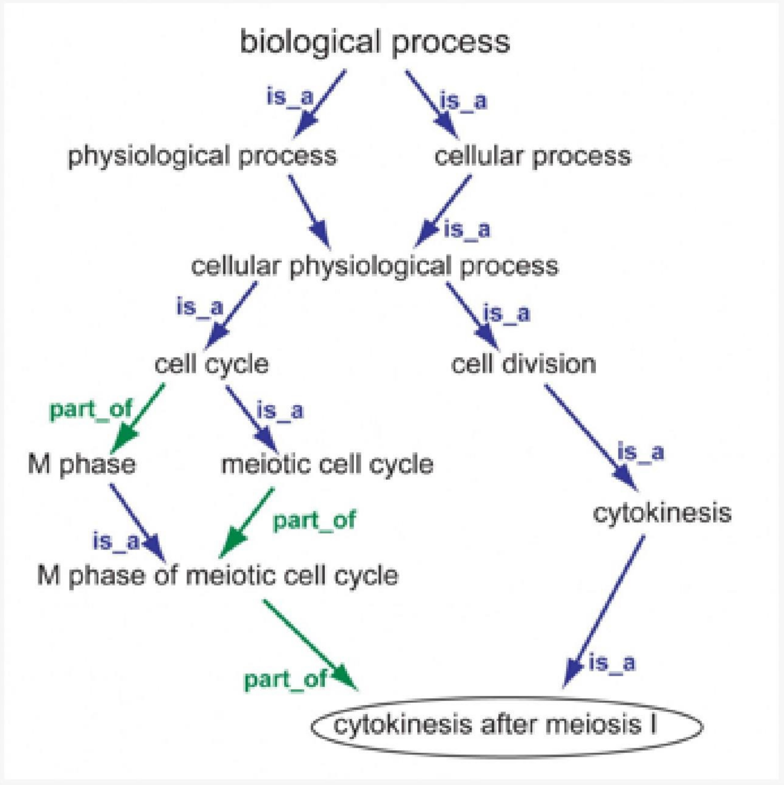

biological process
is_a
is_a
physiological process
cellular process
is_a
cellular physiological process
is_a
is_a
cell cycle
cell division
part_of
is_a
M phase
meiotic cell cycle
is_a
part_of
is_a
cytokinesis
M phase of meiotic cell cycle
part_of
is_a
cytokinesis after meiosis I

Fig. 15.3: A typical, though small, subnetwork of the Gene Ontology's hierarchical network.

the same purposes, they should be linked together. Portrayals of typical heterarchical linkage patterns among natural language concepts are given in Figures 15.5 and 15.6. Just for fun, Figure 15.7 shows one person's attempt to draw a heterarchical graph of the main concepts in one of Douglas Hofstadter's books. Naturally, real concept heterarchies are far more large, complex and tangled than even this one.

In CogPrime, ECAN enforces heterarchy via building SymmetricHebbianLinks, and PLN by building SimilarityLinks, IntensionalSimilarityLinks and ExtensionalSimilarityLinks. Furthermore, these various link types reinforce each other. PLN control is guided by importance spreading, which follows Hebbian links, so that a heterarchical Hebbian network tends to cause PLN to explore the formation of links following the same paths as the heterarchical HebbianLinks. And importance can spread along logical links as well as explicit Hebbian links, so that the existence of a heterarchical logical network will tend to cause the formation of additional heterarchical Hebbian links. Heterarchy reinforces itself in "autopoietic attractor" style even more simply and directly than heterarchy.

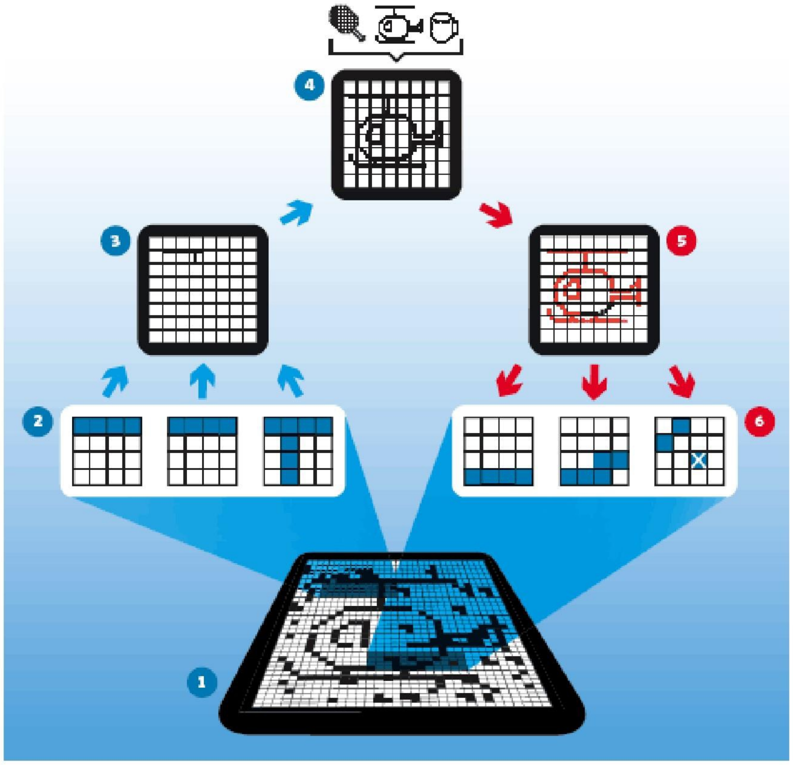


Fig. 15.4: Small-scale portrayal of a portion of the spatiotemporal hierarchy in Jeff Hawkins' Hierarchical Temporal Memory architecture.

15.3.3 Dual Networks

Finally, if both hierarchical and heterarchical structures exist in an Atomspace, then both ECAN and PLN will naturally blend them together, because hierarchical and heterarchical links will feed into their link-creation processes and naturally be combined together to form new links. This will tend to produce a structure called a *dual network*, in which a hierarchy exists, along with a rich network of heterarchical links joining nodes in the hierarchy, with a particular density of links between nodes on the same hierarchical level. The dual network structure will emerge without any explicit engineering oriented toward it, simply via the existence of hierarchical and heterarchical networks, and the propensity of ECAN and PLN to be guided by both the hierarchical and heterarchical networks. The existence of a natural dual network structure in both linguistic and sensorimotor data will help the formation process along, and then creative cognition will enrich the dual network yet further than is directly necessitated by the external world.

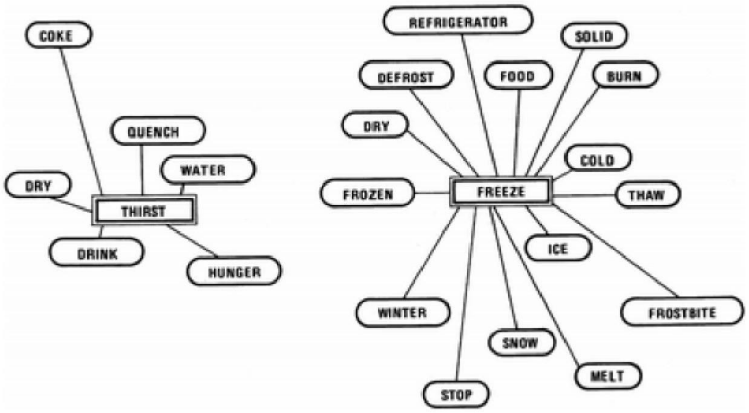

COKE
QUENCH
WATER
DRY
THIRST
DRINK
HUNGER
REFRIGERATOR
SOLID
DEFROST
FOOD
BURN
DRY
COLD
FROZEN
FREEZE
THAW
ICE
WINTER
FROSTBITE
SNOW
MELT
STOP

Fig. 15.5: Portions of a conceptual heterarchy centered on specific concepts.

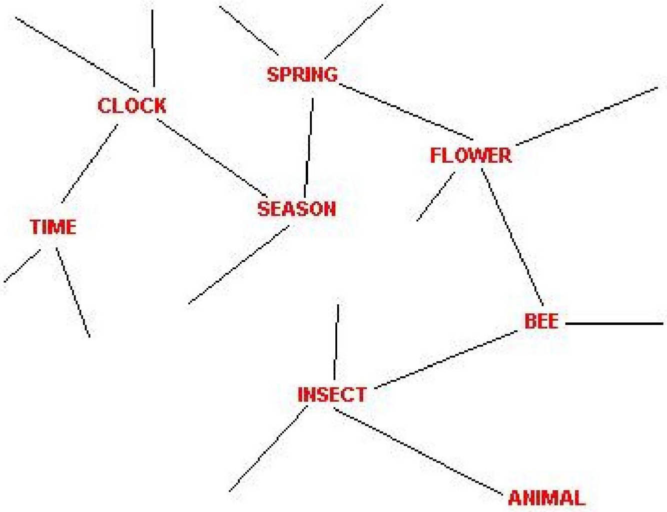

SPRING
CLOCK
FLOWER
TIME
SEASON
BEE
INSECT
ANIMAL

Fig. 15.6: A portion of a conceptual heterarchy, showing the "dangling links" leading this portion to the rest of the heterarchy.

A rigorous mathematical analysis of the formation of hierarchical, heterarchical and dual networks in CogPrime systems has not yet been undertaken, and would certainly be an interesting enterprise. Similar to the theory of small world networks, there is ample ground here for both theorem-proving and heuristic experimentation. However, the qualitative points made here are sufficiently well-grounded in intuition and experience to be of some use guiding our

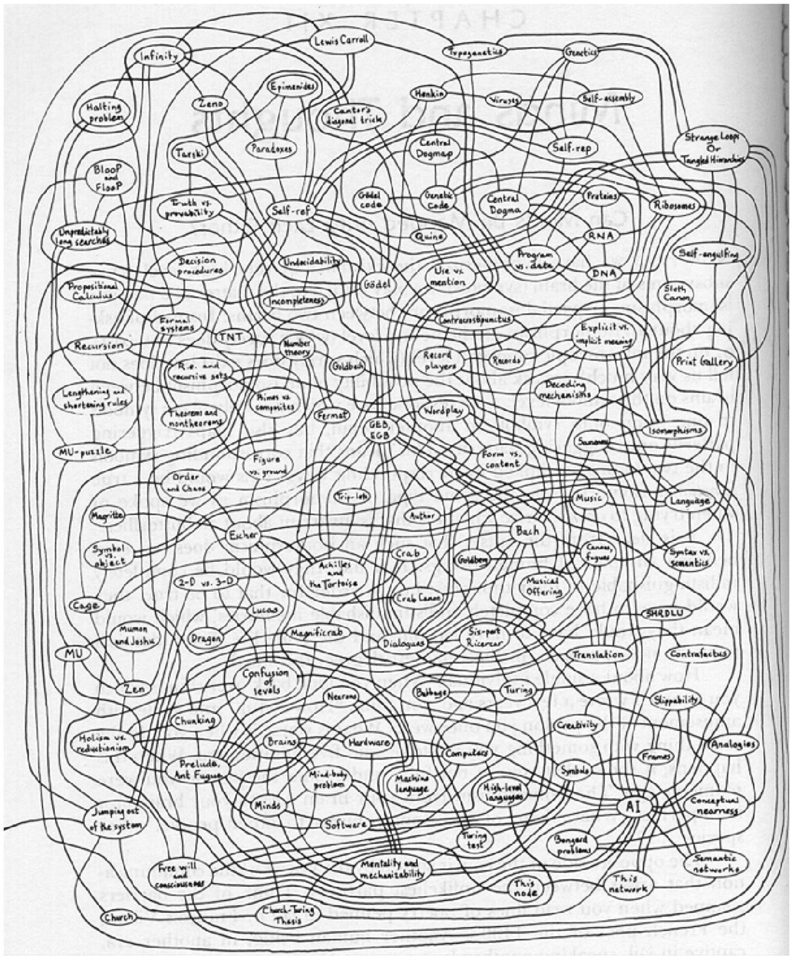

Lewis Carroll
Typogenetics
Genetics
Infinity
Epimenides
Henkin
Viruses
Self-assembly
Halting problem
Zeno
Cantor's diagonal trick
Tarski
Paradoxes
Central Dogmap
Self-rep
Strange Loops Or Tangled Hierarchies
BlooP and FlooP
Truth vs. provability
Self-ref
Gödel code
Genetic code
Central Dogma
Proteins
Ribosomes
Unpredictably long searches
Quine
RNA
Decision procedures
Undecidability
Program vs. data
Self-engulfing
Gödel
Use vs. mention
DNA
Propositional Calculus
Incompleteness
Sloth Canon
Contracrostipunctus
Formal systems
TNT
Explicit vs. implicit meaning
Recursion
Number theory
Record players
Records
Print Gallery
R.e. and recursive sets
Goldbach
Decoding mechanisms
Lengthening and shortening rules
Primes vs. composites
Theorems and nontheorems
Fermat
GEB, EGB
Wordplay
Isomorphisms
MU-puzzle
Form vs. content
Figure vs. ground
Order and Chaos
Trip-lets
Music
Language
Magritte
Author
Bach
Escher
Symbol vs. object
Crab
Goldberg
Canons, fugues
Syntax vs. semantics
Achilles and the Tortoise
Musical Offering
2-D vs. 3-D
Crab Canon
SHRDLU
Cage
Lucas
Mumon and Joshu
Dragon
Magnificrab
Dialogues
Six-part Ricercar
Translation
Contrafactus
MU
Confusion of levels
Zen
Neurons
Babbage
Turing
Slippability
Chunking
Creativity
Holism vs. reductionism
Brains
Hardware
Computers
Analogies
Frames
Prelude, Ant Fugue
Mind-body problem
Machine language
Symbols
High-level languages
Minds
AI
Conceptual nearness
Jumping out of the system
Software
Turing test
Bongard problems
Semantic networks
Free will and consciousness
Mentality and mechanizability
This node
This network
Church
Church-Turing Thesis

Fig. 15.7: A fanciful evocation of part of a reader's conceptual heterarchy related to Douglas Hofstadter's writings.

ongoing work. One of the nice things about emergent network structures is that they are relatively straightforward to observe in an evolving, learning AGI system, via visualization and inspection of structures such at the Atomspace.

Section V

A Path to Human-Level AGI

Chapter 16

AGI Preschool

Co-authored with Stephan Vladimir Bugaj

16.1 Introduction

In conversations with government funding sources or narrow AI researchers about AGI work, one of the topics that comes up most often is that of "evaluation and metrics" – i.e., AGI intelligence testing. We actually prefer to separate this into two topics: environments and methods for careful qualitative evaluation of AGI systems, versus metrics for precise measurement of AGI systems. The difficulty of formulating bulletproof metrics for partial progress toward advanced AGI has become evident throughout the field, and in Chapter 8 we have elaborated one plausible explanation for this phenomenon, the "trickiness" of cognitive synergy. [LWML09], summarizing a workshop on "Evaluation and Metrics for Human-Level AI" held in 2008, discusses some of the general difficulties involved in this type of assessment, and some requirements that any viable approach must fulfill. On the other hand, the lack of appropriate methods for careful qualitative evaluation of AGI systems has been much less discussed, but we consider it actually a more important issue – as well as an easier (though not easy) one to solve.

We haven't actually found the lack of quantitative intelligence metrics to be a major obstacle in our practical AGI work so far. Our OpenCogPrime implementation lags far behind the CogPrime design as articulated in Part 2 of this book, and according to the theory underlying CogPrime, the more interesting behaviors and dynamics of the system will occur only when all the parts of the system have been engineered to a reasonable level of completion and integrated together. So, the lack of a great set of metrics for evaluating the intelligence of our partially-built system hasn't impaired too much. Testing the intelligence of the current OpenCogPrime system is a bit like testing the flight capability of a partly-built airplane that only has stubs for wings, lacks tail-fins, has a much less efficient engine than the one that's been designed for use in the first "real" version of the airplane, etc. There may be something to be learned from such preliminary tests, but making them highly rigorous isn't a great use of effort, compared to working on finishing implementing the design according to the underlying theory.

On the other hand, the problem of what environments and methods to use to qualitatively evaluate and study AGI progress, has been considerably more vexing to us in practice, as we've proceeded in our work on implementing and testing OpenCogPrime and developing the CogPrime theory. When developing a complex system, it's nearly always valuable to see what this system does in some fairly rich, complex situations, in order to gain a better intuitive understanding of the parts and how they work together. In the context of human-level AGI, the theoretically best way to do this would be to embody one's AGI system in a humanlike body and set it loose in the everyday human world; but of course, this isn't feasible given the current state of development of robotics technology. So one must seek approximations. Toward this end we have embodied OpenCogPrime in non-player characters in video game style virtual worlds, and carried out preliminary experiments embodying OpenCogPrime in humanoid robots. These are reasonably good options but they have limitations and lead to subtle choices: what kind of game characters and game worlds, what kind of robot environments, etc.?

One conclusion we have come to, based largely on the considerations in Chapter 11 on development and Chapter 9 on the importance of environment, is that it may make sense to embed early-stage proto-AGI and AGI systems in environments reminiscent of those used for teaching young human children. In this chapter we will explore this approach in some detail: emulation, in either physical reality or an multiuser online virtual world, of an environment similar to preschools used in early human childhood education. Complete specification of an "AGI Preschool" would require much more than a brief chapter; our goal here is to sketch the idea in broad outline, and give a few examples of the types of opportunities such an environment would afford for instruction, spontaneous learning and formal and informal evaluation of certain sorts of early-stage AGI systems.

The material in this chapter will pop up fairly often later in the book. The AGI Preschool context will serve, throughout the following chapters, as a source of concrete examples of the various algorithms and structures. But it's not proposed merely as an expository tool; we are making the very serious proposal that sending AGI systems to a virtual or robotic preschool is an excellent way – perhaps the best way – to foster the development of human-level human-like AGI.

16.1.1 Contrast to Standard AI Evaluation Methodologies

The reader steeped in the current AI literature may wonder why it's necessary to introduce a new methodology and environment for evaluating AGI systems. There are already very many different ways of evaluating AI systems out there ... do we really need another?

Certainly, the AI field has inspired many competitions, each of which tests some particular type or aspect of intelligent behavior. Examples include robot competitions, tournaments of computer chess, poker, backgammon and so forth at computer olympiads, trading-agent competition, language and reasoning competitions like the Pascal Textual Entailment Challenge, and so on. In addition to these, there are many standard domains and problems used in the AI literature that are meant to capture the essential difficulties in a certain class of learning problems: standard datasets for face recognition, text parsing, supervised classification, theorem-proving, question-answering and so forth.

However, the value of these sorts of tests for AGI is predicated on the hypothesis that the degree of success of an AI program at carrying out some domain-specific task, is correlated with the potential of that program for being developed into a robust AGI program with broad intelligence. If humanlike AGI and problem-area-specific "narrow AI" are in fact very different sorts of pursuits requiring very different principles, as we suspect, then these tests are not strongly relevant to the AGI problem.

There are also some standard evaluation paradigms aimed at AI going beyond specific tasks. For instance, there is a literature on "multitask learning" and "transfer learning," where the goal for an AI is to learn one task quicker given another task solved previously [Car97, TM95, BDS03, TS07, RZDK05]. This is one of the capabilities an AI agent will need to simultaneously learn different types of tasks as proposed in the Preschool scenario given here. And there is a literature on "shaping," where the idea is to build up the capability of an AI by training it on progressively more difficult versions of the same tasks [LD03]. Again, this is one sort of capability an AI will need to possess if it is to move up some type of curriculum, such as a school curriculum.

While we applaud the work done on multitask learning and shaping, we feel that exploring these processes using mathematical abstractions, or in the domain of various narrowly-proscribed machine-learning or robotics test problems, may not adequately address the problem of AGI. The problem is that generalization among tasks, or from simpler to more difficult versions of the same task, is a process whose nature may depend strongly on the overall nature of the set of tasks and task-versions involved. Real-world tasks have a subtlety of interconnectedness and developmental course that is not captured in current mathematical learning frameworks nor standard AI test problems.

To put it mathematically, we suggest that the universe of real-world human tasks has a host of "special statistical properties" that have implications regarding what sorts of AI programs will be most suitable; and that, while exploring and formalizing the nature of these statistical properties is important, an easier and more reliable approach to AGI testing is to create a testing environment that embodies these properties implicitly, via its being an emulation of the cognitively meaningful aspects of the real-world human learning environment.

One way to see this point vividly is to contrast the current proposal with the "General Game Player" AI competition, in which AIs seek to learn to play games based on formal descriptions of the rules.[1]. Clearly doing GGP well requires powerful AGI; and doing GGP even mediocrely probably requires robust multitask learning and shaping. But we suspect GGP is far inferior to AGI Preschool as an approach to testing early-stage AI programs aimed at roughly humanlike intelligence. This is because, unlike the tasks involved in AI Preschool, the tasks involved in doing simple instances of GGP seem to have little relationship to humanlike intelligence or real-world human tasks.

16.2 Elements of Preschool Design

What we mean by an "AGI Preschool" is simply a porting to the AGI domain of the essential aspects of human preschools. While there is significant variance among preschools there are also strong commonalities, grounded in educational theory and experience. We will briefly discuss both the physical design and educational curriculum of the typical human preschool, and which aspects transfer effectively to the AGI context.

On the physical side, the key notion in modern preschool design is the "learning center," an area designed and outfitted with appropriate materials for teaching a specific skill. Learning centers are designed to encourage learning by doing, which greatly facilitates learning processes based on reinforcement, imitation and correction (see Chapter 31 of Part 2 for a detailed discussion of the value of this combination); and also to provide multiple techniques for teaching the same skills, to accommodate different learning styles and prevent over-fitting and overspecialization in the learning of new skills.

---

[1] `http://games.stanford.edu/`

Centers are also designed to cross-develop related skills. A “manipulatives center,” for example, provides physical objects such as drawing implements, toys and puzzles, to facilitate development of motor manipulation, visual discrimination, and (through sequencing and classification games) basic logical reasoning. A “dramatics center,” on the other hand, cross-trains interpersonal and empathetic skills along with bodily-kinesthetic, linguistic, and musical skills. Other centers, such as art, reading, writing, science and math centers are also designed to train not just one area, but to center around a primary intelligence type while also cross-developing related areas. For specific examples of the learning centers associated with particular contemporary preschools, see [Nie98].

In many progressive, student-centered preschools, students are left largely to their own devices to move from one center to another throughout the preschool room. Generally, each center will be staffed by an instructor at some points in the day but not others, providing a variety of learning experiences. At some preschools students will be strongly encouraged to distribute their time relatively evenly among the different learning centers, or to focus on those learning centers corresponding to their particular strengths and/or weaknesses.

To imitate the general character of a human preschool, one would create several centers in a robot lab or virtual world. The precise architecture will best be adapted via experience but initial centers would likely be:

- **a blocks center**: a table with blocks on it
- **a language center**: a circle of chairs, intended for people to sit around and talk with the robot
- **a manipulatives center**: with a variety of different objects of different shapes and sizes, intended to teach visual and motor skills
- **a ball play center**: where balls are kept in chests and there is space for the robot to kick the balls around
- **a dramatics center**: where the robot can observe and enact various movements

16.3 Elements of Preschool Curriculum

While preschool curricula vary considerably based on educational philosophy and regional and cultural factors, there is a great deal of common, shared wisdom regarding the most useful topics and methods for preschool teaching. Guided experiential learning in diverse environments and using varied materials is generally agreed upon as being an optimal methodology to reach a wide variety of learning types and capabilities. Hands-on learning provides grounding in specifics, where as a diversity of approaches allows for generalization.

Core knowledge domains are also relatively consistent, even across various philosophies and regions. Language, movement and coordination, autonomous judgment, social skills, work habits, temporal orientation, spatial orientation, mathematics, science, music, visual arts, and dramatics are universal areas of learning which all early childhood learning touches upon. The particulars of these skills may vary, but all human children are taught to function in these domains. The level of competency developed may vary, but general domain knowledge is provided. For example, most kids won’t be the next Maria Callas, Ravi Shankar or Gene Ween, but nearly all learn to hear, understand and appreciate music.

Tables 16.1 - 16.3 review the key capabilities taught in preschools, and identify the most important specific skills that need to be evaluated in the context of each capability. This table was assembled via surveying the curricula from a number of currently existing preschools employing different methodologies both based on formal academic cognitive theories [Sch07] and more pragmatic approaches, such as: Montessori [Mon12], Waldorf [SS03b], Brain Gym (`www.braingym.org`) and Core Knowledge (`www.coreknowledge.org`).

Type of Capability	*Specific Skills to be Evaluated*
Story Understanding	• Understanding narrative sequence • Understanding character development • Dramatize a story • Predict what comes next in a story
Linguistic	• Give simple descriptions of events • Describe similarities and differences • Describe objects and their functions
Linguistic / Spatial-Visual	Interpreting pictures
Linguistic / Social	• Asking questions appropriately • Answering questions appropriately • Talk about own discoveries • Initiate conversations • Settle disagreements • Verbally express empathy • Ask for help • Follow directions
Linguistic / Scientific	• Provide possible explanations for events or phenomena • Carefully describe observations • Draw conclusions from observations

Table 16.1: Categories of Preschool Curriculum, Part 1

16.3.1 Preschool in the Light of Intelligence Theory

Comparing Table 16.1 to Gardner's Multiple Intelligences (MI) framework briefly reviewed in Chapter 2, the high degree of harmony is obvious, and is borne out by more detailed analysis. Preschool curriculum as standardly practiced is very well attuned to MI, and naturally covers all the bases that Gardner identifies as important. And this is not at all surprising since one of Gardner's key motivations in articulating MI theory was the pragmatics of educating humans with diverse strengths and weaknesses.

Regarding intelligence as "the ability to achieve complex goals in complex environments," it is apparent that preschools are specifically designed to pack a large variety of different micro-

Type of Capability	*Specific Skills to be Evaluated*
Logical-Mathematical	• Categorizing • Sorting • Arithmetic • Performing simple "proto-scientific experiments"
Nonverbal Communication	• Communicating via gesture • Dramatizing situations • Dramatizing needs, wants • Express empathy
Spatial-Visual	• Visual patterning • Self-expression through drawing • Navigate
Objective	• Assembling objects • Disassembling objects • Measurement • Symmetry • Similarity between structures (e.g. block structures and real ones)

Table 16.2: Categories of Preschool Curriculum, Part 2

Type of Capability	*Specific Skills to be Evaluated*
Interpersonal	• Cooperation • Display appropriate behavior in various settings • Clean up belongings • Share supplies
Emotional	• Delay gratification • Control emotional reactions • Complete projects

Table 16.3: Categories of Preschool Curriculum, Part 3

environments (the learning centers) into a single room, and to present a variety of different tasks in each environment. The environments constituted by preschool learning centers are designed as microcosms of the most important aspects of the environments faced by humans in their everyday lives.

16.4 Task-Based Assessment in AGI Preschool

Professional pedagogues such as [CM07] discuss evaluation of early childhood learning as intended to assess both specific curriculum content knowledge as well as the child's learning process. It should be as unobtrusive as possible, so that it just seems like another engaging activity, and the results used to tailor the teaching regimen to use different techniques to address weaknesses and reinforce strengths.

For example, with group building of a model car, students are tested on a variety of skills: procedural understanding, visual acuity, motor acuity, creative problem solving, interpersonal communications, empathy, patience, manners, and so on. With this kind of complex, yet engaging, activity as a metric the teacher can see how each student approaches the process of understanding each subtask, and subsequently guide each student's focus differently depending on strengths and weaknesses.

In Tables 16.4 and 16.5 we describe some particular tasks that AGIs may be meaningfully assigned in the context of a general AGI Preschool design and curriculum as described above. Of course, this is a very partial list, and is intended as evocative rather than comprehensive.

Any one of these tasks can be turned into a rigorous quantitative test, thus allowing the precise comparison of different AGI systems' capabilities; but we have chosen not to emphasize this point here, partly for space reasons and partly for philosophical ones. In some contexts the quantitative comparison of different systems may be the right thing to do, but as discussed in Chapter 17 there are also risks associated with this approach, including the emergence of an overly metrics-focused "bakeoff mentality" among system developers, and overfitting of AI abilities to test taking. What is most important is the isolation of specific tasks on which different systems may be experientially trained and then qualitatively assessed and compared, rather than the evaluation of quantitative metrics.

Task-oriented testing allows for feedback on applications of general pedagogical principles to real-world, embodied activities. This allows for iterative refinement based learning (shaping), and cross development of knowledge acquisition and application (multitask learning). It also helps militate against both cheating, and over-fitting, as teachers can make ad-hoc modifications to the tests to determine if this is happening and correct for it if necessary.

E.g., consider a linguistic task in which the AGI is required to formulate a set of instructions encapsulating a given behavior (which may include components that are physical, social, linguistic, etc.). Note that although this is presented as centrally a linguistic task, it actually involves a diverse set of competencies since the behavior to be described may encompass multiple real-world aspects.

To turn this task into a more thorough test one might involve a number of human teachers and a number of human students. Before the test, an ensemble of copies of the AGI would be created, with identical knowledge state. Each copy would interact with a different human teacher, who would demonstrate to it a certain behavior. After testing the AGI on its own knowledge of the material, the teacher would then inform the AGI that it will then be tested on its ability to verbally describe this behavior to another. Then, the teacher goes away and the copy interacts with a series of students, attempting to convey to the students the instructions given by the teacher.

The teacher can thereby assess both the AGI's understanding of the material, and the ability to explain it to the other students. This separates out assessment of understanding from assessment of ability to communicate understanding, attempting to avoid conflation of one with the other. The design of the training and testing needs to account for potential

Intelligence Type	**Test**
Linguistic	• write a set of instructions • speak on a subject • edit a written piece or work • write a speech • commentate on an event • apply positive or negative 'spin' to astory
Logical-Mathematical	• perform arithmetic calculations • create a process to measure something • analyse how a machine works • create a process • devise a strategy to achieve an aim • assess the value of a proposition
Musical	• perform a musical piece • sing a song • review a musical work • coach someone to play a musical instrument
Bodily-Kinesthetic	• juggle • demonstrate a sports technique • flip a beer-mat • create a mime to explain something • toss a pancake • fly a kite

Table 16.4: Prototypical preschool intelligence assessment tasks, Part 1

This testing protocol abstracts away from the particularities of any one teacher or student, and focuses on effectiveness of communication in a human context rather than according to formalized criteria. This is very much in the spirit of how assessment takes place in human preschools (with the exception of the copying aspect): formal exams are rarely given in preschool, but pragmatic, socially-embedded assessments are regularly made.

By including the copying aspect, more rigorous statistical assessments can be made regarding efficacy of different approaches for a given AGI design, independent of past teaching experiences. The multiple copies may, depending on the AGI system design, then be able to be reintegrated, and further "learning" be done by higher-order cognitive systems in the AGI that integrate the disparate experiences of the multiple copies.

This kind of parallel learning is different from both sequential learning that humans do, and parallel presences of a single copy of an AGI (such as in multiple chat rooms type experiments). All three approaches are worthy of study, to determine under what circumstances, and with which AGI designs, one is more successful than another.

It is also worth observing how this test could be tweaked to yield a test of generalization ability. After passing the above, the AGI could then be given a description of a new task

Intelligence Type	**Test**
Spatial-Visual	• design a costume • interpret a painting • create a room layout • create a corporate logo • design a building • pack a suitcase or the trunk of a car
Interpersonal	• interpret moods from facial expressions • demonstrate feelings through body language • affect the feelings of others in a planned way • coach or counsel another

Table 16.5: Prototypical preschool intelligence assessment tasks, Part 2

(acquisition), and asked to explain the new one (variation). And, part of the training behavior might be carried out unobserved by the AGI, thus requiring the AGI to infer the omitted parts of the task it needs to describe.

Another popular form of early childhood testing is puzzle block games. These kinds of games can be used to assess a variety of important cognitive skills, and to do so in a fun way that not only examines but also encourages creativity and flexible thinking. Types of games include pattern matching games in which students replicate patterns described visually or verbally, pattern creation games in which students create new patterns guided by visually or verbally described principles, creative interpretation of patterns in which students find meaning in the forms, and free-form creation. Such games may be individual or cooperative.

Cross training and assessment of a variety of skills occurs with pattern block games: for example, interpretation of visual or linguistic instructions, logical procedure and pattern following, categorizing, sorting, general problem solving, creative interpretation, experimentation, and kinematic acuity. By making the games cooperative, various interpersonal skills involving communication and cooperation are also added to the mix.

The puzzle block context bring up some general observations about the role of kinematic and visuospatial intelligence in the AGI Preschool. Outside of robotics and computer vision, AI research has often downplayed these sorts of intelligence (though, admittedly, this is changing in recent years, e.g. with increasing research focus on diagrammatic reasoning). But these abilities are not only necessary to navigate real (or virtual) spatial environments. They are also important components of a coherent, conceptually well-formed understanding of the world in which the student is embodied. Integrative training and assessment of both rigorous cognitive abilities generally most associated with both AI and "proper schooling" (such as linguistic and logical skills) along with kinematic and aesthetic/sensory abilities is essential to the development of an intelligence that can successfully both operate in and sensibly communicate about the real world in a roughly humanlike manner. Whether or not an AGI is targeted to interpret physical-world spatial data and perform tasks via robotics, in order to communicate ideas about a vast array of topics of interest to any intelligence in this world, an AGI must develop aspects of intelligence other than logical and linguistic cognition.

16.5 Beyond Preschool

Once an AGI passes preschool, what are the next steps? There is still a long way to go, from preschool to an AGI system that is capable of, say, passing the Turing Test or serving as an effective artificial scientist.

Our suggestion is to extend the school metaphor further, and make use of existing curricula for higher levels of virtual education: grade school, secondary school, and all levels of post-secondary education. If an AGI can pass online primary and secondary schools such as e-tutor.com, and go on to earn an online degree from an accredited university, then clearly said AGI has successfully achieved "human level, roughly humanlike AGI." This sort of testing is interesting not only because it allows assessment of stages intermediate between preschool and adult, but also because it tests humanlike intelligence without requiring precise imitation of human behavior.

If an AI can get a BA degree at an accredited university, via online coursework (assuming for simplicity courses where no voice interaction is needed), then we should consider that AI to have human-level intelligence. University coursework spans multiple disciplines, and the details of the homework assignments and exams are not known in advance, so like a human student the AGI team can't cheat.

In addition to the core coursework, a schooling approach also tests basic social interaction and natural language communication, ability to do online research, and general problem solving ability. However, there is no rigid requirement to be strictly humanlike in order to pass university classes.

Most of our concrete examples in the following chapters will pertain to the preschool context, because it's simple to understand, and because we feel that getting to the "AGI preschool student" level is going to be the largest leap. Once that level is obtained, moving further will likely be difficult also, but we suspect it will be more a matter of steady incremental improvements – whereas the achievement of preschool-level functionality will be a large leap from the current situation.

16.6 Issues with Virtual Preschool Engineering

As noted above there are two broad approaches to realizing the "AGI Preschool" idea: using the AGI to control a physical robot and then crafting a preschool environment suitable to the robot's sensors and actuators; or, using the AGI to control a virtual agent in an appropriately rich virtual-world preschool. The robotic approach is harder from an AI perspective (as one must deal with problems of sensation and actuation), but easier from an environment-construction perspective. In the virtual world case, one quickly runs up against the current limitations of virtual world technologies, which have been designed mainly for entertainment or social-networking purposes, not with the requirements of AGI systems in mind.

In Chapter 9 we discussed the general requirements that an environment should possess to be supportive of humanlike intelligence. Referring back to that list, it's clear that current virtual worlds are fairly strong on multimodal communication, and fairly weak on naive physics. More concretely, if one wants a virtual world so that

1. one could carry out all the standard cognitive development experiments described in developmental psychology books
2. one could implement intuitively reasonable versions of all the standard activities in all the standard learning stations in a contemporary preschool

then current virtual world technologies appear not to suffice.

As reviewed above, typical preschool activities include for instance building with blocks, playing with clay, looking in a group at a picture book and hearing it read aloud, mixing ingredients together, rolling/throwing/catching balls, playing games like tag, hide-and-seek, Simon Says or Follow the Leader, measuring objects, cutting paper into different shapes, drawing and coloring, etc.

And, as typical, not necessarily representative examples of tasks psychologists use to measure cognitive development (drawn mainly from the Piagetan tradition, without implying any assertion that this is the only tradition worth pursuing), consider the following:

1. Which row has more circles- A or B? A: O O O O O, B: OOOOO
2. If Mike is taller than Jim, and Jim is shorter than Dan, then who is the shortest? Who is the tallest?
3. Which is heavier- a pound of feathers or a pound of rocks?
4. Eight ounces of water is poured into a glass that looks like the fat glass in Figure 2 16.1 and then the same amount is poured into a glass that looks like the tall glass in Figure 16.2 . Which glass has more water?
5. A lump of clay is rolled into a snake. All the clay is used to make the snake. Which has more clay in it – the lump or the snake?
6. There are two dolls in a room, Sally and Ann, each of which has her own box, with a marble hidden inside. Sally goes out for a minute, leaving her box behind; and Ann decides to play a trick on Sally: she opens Sally's box, removes the marble, hiding it in her own box. Sally returns, unaware of what happened. Where will Sally would look for her marble?
7. Consider this rule about a set of cards that have letters on one side and numbers on the other: "If a card has a vowel on one side, then it has an even number on the other side." If you have 4 cards labeled "E K 4 7", which cards do you need to turn over to tell if this rule is actually true?
8. Design an experiment to figure out how to make a pendulum that swings more slowly versus less slowly

What we see from this ad hoc, partial list is that a lot of naive physics *is* required to make an even vaguely realistic preschool. A lot of preschool education is about the intersection between abstract cognition and naive physics. A more careful review of the various tasks involved in preschool education bears out this conclusion.

With this in mind, in this section we will briefly describe an approach to extending current virtual world technologies that appears to allow the construction of a reasonably rich and realistic AGI preschool environment, without requiring anywhere near a complete simulation of realistic physics.

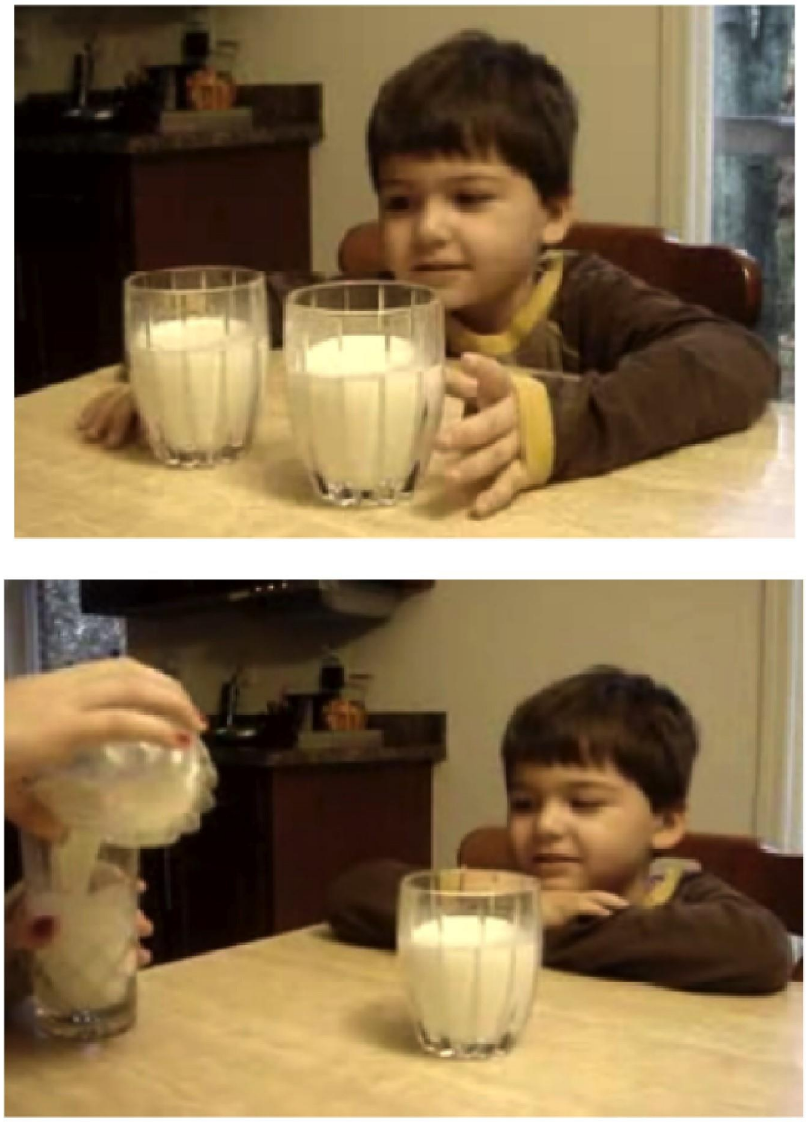


Fig. 16.1: Part 1 of a Piagetan conservation of volume experiment: a child observes that two glasses obviously have the same amount of milk in them, and then sees the content of one of the glasses poured into a different-shaped glass.

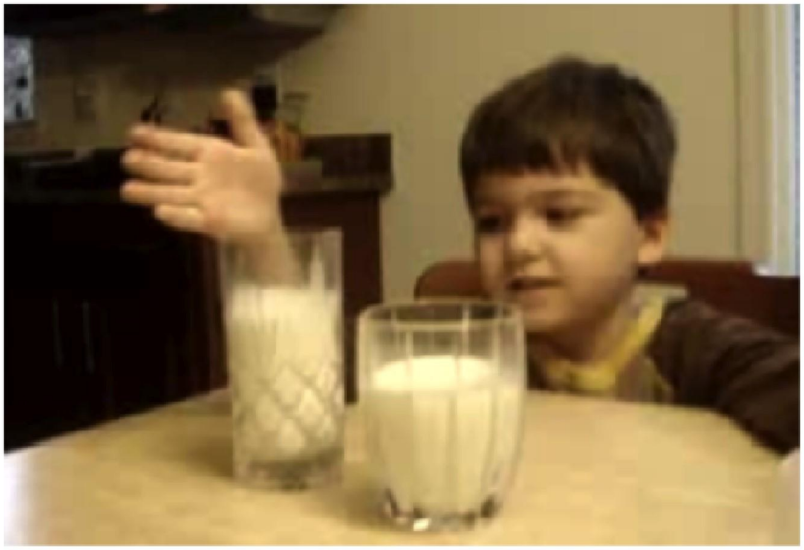


Fig. 16.2: Part 2 of a Piagetan conservation of volume experiment: a child observes two different-shaped glasses, which (depending on the level of his cognition), he may be able to infer have the same amount of milk in them, due to the events depicted in Figure 16.1.

16.6.1 Integrating Virtual Worlds with Robot Simulators

One glaring deficit in current virtual world platforms is the lack of flexibility in terms of tool use. In most of these systems today, an avatar can pick up or utilize an object, or two objects can interact, only in specific, pre-programmed ways. For instance, an avatar might be able to pick up a virtual screwdriver only by the handle, rather than by pinching the blade betwen its fingers. This places severe limits on creative use of tools, which is absolutely critical in a preschool context. The solution to this problem is clear: adapt existing generalized physics engines to mediate avatar-object and object-object interactions. This would require more computation than current approaches, but not more than is feasible in a research context.

One way to achieve this goal would be to integrate a robot simulator with a virtual world or game engine, for instance to modify the OpenSim (`opensimulator.org`) virtual world to use the Gazebo (`playerstage.sourceforge.net`) robot simulator in place of its current physics engine. While tractable, such a project would require considerable software engineering effort.

16.6.2 BlocksNBeads World

Another glaring deficit in current virtual world platforms is their inability to model physical phenomena besides rigid objects with any sophistication. In this section we propose a potential solution to this issue: a novel class of virtual worlds called BlocksNBeadsWorld, consisting of the following aspects:

1. 3D blocks of various shapes and sizes and frictional coefficients, that can be stacked
2. Adhesive that can be used to stick blocks together, and that comes in two types, one of which can be removed by an adhesive-removing substance, one of which cannot (though its bonds can be broken via sufficient application of force)
3. Spherical beads, each of which has intrinsic unchangeable adhesion properties defined according to a particular, simple "adhesion logic"
4. Each block, and each bead, may be associated with multidimensional quantities representing its taste and smell; and may be associated with a set of sounds that are made when it is impacted with various forces at various positions on its surface

Interaction between blocks and beads is to be calculated according to standard Newtonian physics, which would be compute-intensive in the case of a large number of beads, but tractable using distributed processing. For instance if 10K beads were used to cover a humanoid agent's face, this would provide a fairly wide diversity of facial expressions; and if 10K beads were used to form a blanket laid on a bed, this would provide a significant amount of flexibility in terms of rippling, folding and so forth. Yet, this order of magnitude of interactions is very small compared to what is done in contemporary simulations of fluid dynamics or, say, quantum chromodynamics.

One key aspect of the spherical beads is that they can be used to create a variety of rigid or flexible surfaces, which may exist on their own or be attached to blocks-based constructs. The specific inter-bead adhesion properties of the beads could be defined in various ways, and will surely need to be refined via experimentation, but a simple scheme that seems to make sense is as follows.

Each bead can have its surface tesselated into hexagons (the number of these can be tuned), and within each hexagon it can have two different adhesion coefficients: one for adhesion to other beads, and one for adhesion to blocks. The adhesion between two beads along a certain hexagon is then determined by their two adhesion coefficients; and the adhesion between a bead and a block is determined by the adhesion coefficient of the bead, and the adhesion coefficient of the adhesive applied to the block. A distinction must be drawn between rigid and flexible adhesion: rigid adhesion sticks a bead to something in a way that can't be removed except via breaking it off; whereas flexible adhesion just keeps a bead very close to the thing it's stuck onto. Any two entities may be stuck together either rigidly or flexibly. Sets of beads with flexible adhesion to each other can be used to make entities like strings, blankets or clothes.

Using the above adhesion logic, it seems one could build a wide variety of flexible structures using beads, such as (to give a very partial list):

1. fabrics with various textures, that can be draped over blocks structures,
2. multilayered coatings to be attached to blocks structures, serving (among many other examples) as facial expressions
3. liquid-type substances with varying viscosities, that can be poured between different containers, spilled, spread, etc.
4. strings tyable in knots; rubber bands that can be stretched; etc.

Of course there are various additional features one could add. For instance one could add a special set of rules for vibrating strings, allowing BlocksNBeadsWorld to incorporate the creation of primitive musical instruments. Variations like this could be helpful but aren't necessary for the world to serve its essential purpose.

Note that one does not have true fluid dynamics in BlocksNBeadsWorld, but, it seems that the latter is not necessary to encompass the phenomena covered in cognitive developmental tests or preschool tasks. The tests and tasks that are done with fluids can instead be done with masses of beads. For example, consider the conservation of volume task shown in Figures 16.1 and 16.2 below: it's easy enough to envision this being done with beads rather than milk. Even a few hundred beads is enough to be psychologically perceived as a mass rather than a set of discrete units, and to be manipulated and analyzed as such. And the simplification of not requiring fluid mechanics in one's virtual world is immense.

Next, one can implement equations via which the adhesion coefficients of a bead are determined in part by the adhesion coefficients of nearby beads, or beads that are nearby in certain directions (with direction calculated in local spherical coordinates). This will allow for complex cracking and bending behaviors – not identical to those in the real world, but with similar qualitative characteristics. For example, without this feature one could create paperlike substances that could be cut with scissors – but *with* this feature, one could go further and create woodlike substances that would crack when nails were hammered into them in certain ways, and so forth.

Further refinements are certainly possible also. One could add multidimensional adhesion coefficients, allowing more complex sorts of substances. One could allow beads to vibrate at various frequencies, which would lead to all sorts of complex wave patterns in bead compounds. Etc. In each case, the question to be asked is: what important cognitive abilities are dramatically more easily learnable in the presence of the new feature than in its absence?

The combination of blocks and beads seems ideal for implementing a more flexible and AGI-friendly type of virtual body than is currently used in games and virtual worlds. One can easily envision implementing a body with

1. a skeleton whose bones consist of appropriately shaped blocks
2. joints consisting of beads, flexibly adhered to the bones
3. flesh consisting of beads, flexibly adhered to each other
4. internal "plumbing" consisting of tubes whose walls are beads rigidly adhered to each other, and flexibly adhered to the surrounding flesh (the plumbing could then serve to pass beads through, where slow passage would be ensured by weak adhesion between the walls of the tubes and the beads passing through the tubes)

This sort of body would support rich kinesthesia; and rich, broad analogy-drawing between the internally-experienced body and the externally-experienced world. It would also afford many interesting opportunities for flexible movement control. Virtual animals could be created along with virtual humanoids.

Regarding the extended mind, it seems clear that blocks and beads are adequate for the creation of a variety of different tools. Equipping agents with "glue guns" able to affect the adhesive properties of both blocks and beads would allow a diversity of building activity; and building with masses of beads could become a highly creative activity. Furthermore, beads with appropriately specified adhesion (within the framework outlined above) could be used to form organically growing plant-like substances, based on the general principles used in L-system models of plant growth (Prusinciewicz and Lindenmayer 1991). Structures with only beads would vaguely resemble herbaceous plants; and structures involving both blocks and beads would more resemble woody plants. One could even make organic structures that flourish or otherwise based on the light available to them (without of course trying to simulate the chemistry of photosynthesis).

Some elements of chemistry may be achieved as well, though nowhere near what exists in physical reality. For instance, melting and boiling at least should be doable: assign every bead a temperature, and let solid interbead bonds turn liquid above a certain temperature and disappear completely above some higher temperature. You could even have a simple form of fire. Let fire be an element, whose beads have negative gravitational mass. Beads of fuel elements like wood have a threshold temperature above which they will turn into fire beads, with release of additional heat.[2]

The philosophy underlying these suggested bead dynamics is somewhat comparable to that outlined in Wolfram's book *A New Kind of Science* [Wol02]. There he proposes cellular automata models that emulate the qualitative characteristics of various real-world phenomena, without trying to match real-world data precisely. For instance, some of his cellular automata demonstrate phenomena very similar to turbulent fluid flow, without implementing the Navier-Stokes equations of fluid dynamics or trying to precisely match data from real-world turbulence. Similarly, the beads in BlocksNBeadsWorld are intended to qualitatively demonstrate the real-world phenomena most useful for the development of humanlike embodied intelligence, without trying to precisely emulate the real-world versions of these phenomena.

The above description has been left imprecisely specified on purpose. It would be straightforward to write down a set of equations for the block and bead interactions, but there seems little value in articulating such equations without also writing a simulation involving them and testing the ensuing properties. Due to the complex dynamics of bead interactions, the fine-tuning of the bead physics is likely to involve some tuning based on experimentation, so that any equations written down now would likely be revised based on experimentation anyway. Our goal here has been to outline a certain class of potentially useful environments, rather than to articulate a specific member of this class.

Without the beads, BlocksNBeadsWorld would appear purely as a "Blocks World with Glue" – essentially a substantially upgraded version of the Blocks Worlds frequently used in AI, since first introduced in [Win72]. Certainly a pure "Blocks World with Glue" would have greater simplicity than BlocksNBeadsWorld, and greater richness than standard Blocks World; but this simplicity comes with too many limitations, as shown by consideration of the various naive physics requirements inventoried above. One simply cannot run the full spectrum of humanlike cognitive development experiments, or preschool educational tasks, using blocks and glue alone. One can try to create analogous tasks using only blocks and glue, but this quickly becomes extremely awkward. Whereas in the BlocksNBeadsWorld the capability for this full spectrum of experiments and tasks seems to fall out quite naturally.

What's missing from BlocksNBeadsWorld should be fairly obvious. There isn't really any distinction between a fluid and a powder: there are masses, but the types and properties of the masses are not the same as in the real world, and will surely lack the nuances of real-world fluid dynamics. Chemistry is also missing: processes like cooking and burning, although they can be crudely emulated, will not have the same richness as in the real world. The full complexity of body processes is not there: the body-design method mentioned above is far richer and more adaptive and responsive than current methods of designing virtual bodies in 3DSMax or Maya and importing them into virtual world or game engines, but still drastically simplistic compared to real bodies with their complex chemical signaling systems and couplings with other bodies and the environment. The hypothesis we're making in this section is that these lacunae aren't that important from the point of view of humanlike cognitive development. We suggest that the key features of naive physics and folk psychology enumerated above can be mastered by an AGI in BlocksNBeadsWorld in spite of its limitations, and that – together with an appropriate AGI design – this probably suffices for creating an AGI with the inductive biases constituting humanlike intelligence.

---

[2] Thanks are due to Russell Wallace for the suggestions in this paragraph

To drive this point home more thoroughly, consider three potential virtual world scenarios:

1. A world containing realistic fluid dynamics, where a child can pour water back and forth between two cups of different shapes and sizes, to understand issues such as conservation of volume
2. A world more like today's Second Life, where fluids don't really exist, and things like lakes are simulated via very simple rules, and pouring stuff back and forth between cups doesn't happen unless it's programmed into the cups in a very specialized way
3. A BlocksNBeadsWorld type world, where a child can pour masses of beads back and forth between cups, but not masses of liquid

Our qualitative judgment is that Scenario 3 is going to allow a young AI to gain the same essential insights as Scenario 1, whereas Scenario 2 is just too impoverished. I have explored dozens of similar scenarios regarding different preschool tasks or cognitive development experiments, and come to similar conclusions across the board. Thus, our current view is that something like BlocksNBeadsWorld can serve as an adequate infrastructure for an AGI Preschool, supporting the development of human-level, roughly human-like AGI.

And, if this view turns out to be incorrect, and BlocksNBeadsWorld is revealed as inadequate, then we will very likely still advocate the conceptual approach enunciated above as a guide for designing virtual worlds for AGI. That is, we would suggest to explore the hypothetical failure of BlocksNBeadsWorld via asking two questions:

1. Are there basic naive physics or folk psychology requirements that were missed in creating the specifications, based on which the adequacy of BlocksNBeadsWorld was assessed?
2. Does BlocksNBeadsWorld fail to sufficiently emulate the real world in respect to some of the articulated naive physics or folk psychology requirements?

The answers to these questions would guide the improvement of the world or the design of a better one.

Regarding the practical implementation of BlocksNBeadsWorld, it seems clear that this is within the scope of modern game engine technology, however, it is not something that could be encompassed within an existing game or world engine without significant additions; it would require substantial custom engineering. There exist commodity and open-source physics engines that efficiently carry out Newtonian mechanics calculations; while they might require some tuning and extension to handle BlocksNBeadWorld, the main issue would be achieving adequate speed of physics calculation, which given current technology would need to be done via modifying existing engines to appropriately distribute processing among multiple GPUs.

Finally, an additional avenue that merits mention is the use of BlocksNBeads physics internally within an AGI system, as part of an internal simulation world that allows it to make "mind's eye" estimative simulations of real or hypothetical physical situations. There seems no reason that the same physics software libraries couldn't be used both for the external virtual world that the AGI's body lives in, and for an internal simulation world that the AGI uses as a cognitive tool. In fact, the BlocksNBeads library could be used as an internal cognitive tool by AGI systems controlling physical robots as well. This might require more tuning of the bead dynamics to accord with the dynamics of various real-world systems; but, this tuning would be beneficial for the BlocksNBeadWorld as well.

Chapter 17
A Preschool-Based Roadmap to Advanced AGI

17.1 Introduction

Supposing the CogPrime approach to creating advanced AGI is workable – then what are the right practical steps to follow? The various structures and algorithms outlined in Part 2 of this book should be engineered and software-tested, of course – but that's only part of the study. The AGI system implemented will need to be taught, and it will need to be placed in situations where it can develop an appropriate self-model and other critical internal network structures. The complex structures and algorithms involved will need to be fine-tuned in various ways, based on qualitatively observing the overall system's behavior in various situations. To get all this right without excessive confusion or time-wastage requires a fairly clear *roadmap* for CogPrime development.

In this chapter we'll sketch one particular roadmap for the development of human-level, roughly human-like AGI – which we're not selling as the only one, or even necessarily as the best one. It's just one roadmap that we have thought about a lot, and that we believe has a strong chance of proving effective. Given resources to pursue only one path for AGI development and teaching, this would be our choice, at present. The roadmap outlined here is not restricted to CogPrime in any highly particular ways, but it has been developed largely with CogPrime in mind; those developing other AGI designs could probably use this roadmap just fine, but might end up wanting to make various adjustments based on the strengths and weaknesses of their own approach.

What we mean here by a "roadmap" is, in brief: a sequence of "milestone" tasks, occurring in a small set of common environments or "scenarios," organized so as to lead to a commonly agreed upon set of long-term goals. I.e., what we are after here is a "capability roadmap" – a roadmap laying out a series of capabilities whose achievement seems likely to lead to human-level AGI. Other sorts of roadmaps such as "tools roadmaps" may also be valuable, but are not our concern here.

More precisely, we confront the task of roadmapping by identifying scenarios in which to embed our AGI system, and then "competency areas" in which the AGI system must be evaluated. Then, we envision a roadmap as consisting of a set of one or more task-sets, where each task set is formed from a combination of a scenario with a list of competency areas. To create a task-set one must choose a particular scenario, and then articulate a set of specific tasks, each one addressing one or more of the competency areas. Each task must then get associated with particular performance metrics – quantitative wherever possible, but perhaps qualitative in some cases depending on the nature of the task. Here we give a partial task-set for the "virtual and robot preschool" scenarios discussed in Chapter 16, and a couple example quantitative metrics just to illustrate what is intended; the creation of a fully detailed roadmap based on the ideas outlined here is left for future work.

The train of thought presented in this chapter emerged in part from a series of conversations preceding and during the "AGI Roadmap Workshop" held at the University of Tennessee, Knoxville in October 2008. Some of the ideas also trace back to discussions held during two workshops on "Evaluation and Metrics for Human-Level AI" organized by John Laird and Pat Langley (one in Ann Arbor in late 2008, and one in Tempe in early 2009). Some of the conclusions of the Ann Arbor workshop were recorded in [LWML09]. Inspiration was also obtained from discussion at the "Future of AGI" post-conference workshop of the AGI-09 conference, triggered by Itamar Arel's [ARK09a] presentation on the "AGI Roadmap" theme; and from an earlier article on AGI Roadmapping by [AL09].

However, the focus of the AGI Roadmap Workshop was considerably more general than the present chapter. Here we focus on preschool-type scenarios, whereas at the workshop a number of scenarios were discussed, including the preschool scenarios but also, for example,

- Standardized Tests and School Curricula
- Elementary, Middle and High School Student
- General Videogame Learning
- Wozniak's Coffee Test: go into a random American house and figure out how to make coffee, and do it
- Robot College Student
- General Call Center Respondent

For each of these scenarios, one may generate tasks corresponding to each of the competency areas we will outline below. CogPrime is applicable in all these scenarios, so our choice to focus on preschool scenarios is an additional judgment call beyond those judgment calls required to specify the CogPrime design. The roadmap presented here is a "AGI Preschool Roadmap" and as such is a special case of the broader "AGI Roadmap" outlined at the workshop.

17.2 Measuring Incremental Progress Toward Human-Level AGI

In Chapter 2, we discussed several examples of practical goals that we find to plausibly characterize "human level AGI", e.g.

- *Turing Test*
- *Virtual World Turing Test*
- *Online University Test*
- *Physical University Test*
- *Artificial Scientist Test*

We also discussed our optimism regarding the possibility that in the future AGI may advance beyond the human level, rendering all these goals "early-stage subgoals."

However, in this chapter we will focus our attention on the nearer term. The above goals are ambitious ones, and while one can talk a lot about how to precisely measure their achievement, we don't feel that's the most interesting issue to ponder at present. More critical is to think about how to measure incremental progress. How do you tell when you're 25% or 50% of the way to having an AGI that can pass the Turing Test, or get an online university degree. Fooling 50% of the Turing Test judges is not a good measure of being 50% of the way to passing the Turing Test (that's too easy); and passing 50% of university classes is not a good measure of being 50% of the way to getting an online university degree (it's too hard – if one had an AGI capable of doing that, one would almost surely be very close to achieving the end goal). Measuring incremental progress toward human-level AGI is a subtle thing, and we argue that the best way to do it is to focus on particular scenarios and the achievement of specific competencies therein.

As we argued in Chapter 8 there are some theoretical reasons to doubt the possibility of creating a rigorous objective test for partial progress toward AGI – a test that would be convincing to skeptics, and impossible to "game" via engineering a system specialized to the test. Fortunately, though we don't need a test of this nature for the purposes of assessing our own incremental progress toward advanced AGI, based on our knowledge about our own approach.

Based on the nature of the grand goals articulated above, there seems to be a very natural approach to creating a set of incremental capabilities building toward AGI: *to draw on our copious knowledge about human cognitive development.* This is by no means the only possible path; one can envision alternatives that have nothing to do with human development (and those might also be better suited to non-human AGIs). However, so much detailed knowledge about human development is available – as well as solid knowledge that the human developmental trajectory does lead to human-level AI – that the motivation to draw on human cognitive development is quite strong.

The main problem with the human development inspired approach is that cognitive developmental psychology is not as systematic as it would need to be for AGI to be able to translate it directly into architectural principles and requirements. As noted above, while early thinkers like Piaget and Vygotsky outlined systematic theories of child cognitive development, these are no longer considered fully accurate, and one currently faces a mass of detailed theories of various aspects of cognitive development, but without an unified understanding. Nevertheless we believe it is viable to work from the human-development data and understanding currently available, and craft a workable AGI roadmap therefrom.

With this in mind, what we give next is a fairly comprehensive list of the competencies that we feel AI systems should be expected to display in one or more of these scenarios in order to be considered as full-fledged "human level AGI" systems. These competency areas have been assembled somewhat opportunistically via a review of the cognitive and developmental psychology literature as well as the scope of the current AI field. We are not claiming this as a precise or exhaustive list of the competencies characterizing human-level general intelligence, and will be happy to accept additions to the list, or mergers of existing list items, etc. What we are advocating is not this specific list, but rather the approach of enumerating competency areas, and then generating tasks by combining competency areas with scenarios.

We also give, with each competency, an example task illustrating the competency. The tasks are expressed in the robot preschool context for concreteness, but they all apply to the virtual preschool as well. Of course, these are only examples, and ideally to teach an AGI in a structured way one would like to

- associate several tasks with each competency
- present each task in a graded way, with multiple subtasks of increasing complexity
- associate a quantitative metric with each task

However, the briefer treatment given here should suffice to give a sense for how the competencies manifest themselves practically in the AGI Preschool context.

1. Perception
    - **Vision**: image and scene analysis and understanding
        - *Example task:* When the teacher points to an object in the preschool, the robot should be able to identify the object and (if it's a multi-part object) its major parts. If it can't perform the identification initially, it can approach the object and manipulate it before making its identification.
    - **Hearing**: identifying the sounds associated with common objects; understanding which sounds come from which sources in a noisy environment
        - *Example task:* When the teacher covers the robot's eyes and then makes a noise with an object, the robot should be able to guess what the object is
    - **Touch**: identifying common objects and carrying out common actions using touch alone
        - *Example task:* With its eyes and ears covered, the robot should be able to identify some object by manipulating it; and carry out some simple behaviors (say, putting a block on a table) via touch alone
    - **Crossmodal**: Integrating information from various senses
        - *Example task:* Identifying an object in a noisy, dim environment via combining visual and auditory information
    - **Proprioception**: Sensing and understanding what its body is doing
        - *Example task:* The teacher moves the robot's body into a certain configuration. The robot is asked to restore its body to an ordinary standing position, and then repeat the configuration that the teacher moved it into.
2. Actuation
    - **Physical skills**: manipulating familiar and unfamiliar objects
        - *Example task:* Manipulate blocks based on imitating the teacher: e.g. pile two blocks atop each other, lay three blocks in a row, etc.
    - **Tool use**, including the flexible use of ordinary objects as tools
        - *Example task:* Use a stick to poke a ball out of a corner, where the robot cannot directly reach
    - **Navigation**, including in complex and dynamic environments
        - *Example task:* Find its own way to a named object or person through a crowded room with people walking in it and objects laying on the floor.
3. Memory
    - **Declarative**: noticing, observing and recalling facts about its environment and experience
        - *Example task:* If certain people habitually carry certain objects, the robot should remember this (allowing it to know how to find the objects when the relevant people are present, even much later)
    - **Behavioral**: remembering how to carry out actions
        - *Example task:* If the robot is taught some skill (say, to fetch a ball), it should remember this much later
    - **Episodic**: remembering significant, potentially useful incidents from life history

  - *Example task:* Ask the robot about events that occurred at times when it got particularly much, or particularly little, reward for its actions; it should be able to answer simple questions about these, with significantly more accuracy than about events occurring at random times

4. Learning

   - **Imitation**: Spontaneously adopt new behaviors that it sees others carrying out
     - *Example task:* Learn to build towers of blocks by watching people do it
   - **Reinforcement**: Learn new behaviors from positive and/or negative reinforcement signals, delivered by teachers and/or the environment
     - *Example task:* Learn which box the red ball tends to be kept in, by repeatedly trying to find it and noticing where it is, and getting rewarded when it finds it correctly
   - **Imitation/Reinforcement**
     - *Example task:* Learn to play "fetch", "tag" and "follow the leader" by watching people play it, and getting reinforced on correct behavior
   - **Interactive Verbal Instruction**
     - *Example task:* Learn to build a particular structure of blocks faster based on a combination of imitation, reinforcement and verbal instruction, than by imitation and reinforcement without verbal instruction
   - **Written Media**
     - *Example task:* Learn to build a structure of blocks by looking at a series of diagrams showing the structure in various stages of completion
   - **Learning via Experimentation**
     - *Example task:* Ask the robot to slide blocks down a ramp held at different angles. Then ask it to make a block slide fast, and see if it has learned how to hold the ramp to make a block slide fast.

5. Reasoning

   - **Deduction**, from uncertain premises observed in the world
     - *Example task:* If Ben more often picks up red balls than blue balls, and Ben is given a choice of a red block or blue block to pick up, which is he more likely to pick up?
   - **Induction**, from uncertain premises observed in the world
     - *Example task:* If Ben comes into the lab every weekday morning, then is Ben likely to come to the lab today (a weekday) in the morning?
   - **Abduction**, from uncertain premises observed in the world
     - *Example task:* If women more often give the robot food than men, and then someone of unidentified gender gives the robot food, is this person a man or a woman?
   - **Causal reasoning**, from uncertain premises observed in the world
     - *Example task:* If the robot knows that knocking down Ben's tower of blocks makes him angry, then what will it say when asked if kicking the ball at Ben's tower of blocks will make Ben mad?
   - **Physical reasoning**, based on observed "fuzzy rules" of naive physics
     - *Example task:* Given two balls (one rigid and one compressible) and two tunnels (one significantly wider than the balls, one slightly narrower than the balls), can the robot guess which balls will fit through which tunnels?
   - **Associational reasoning**, based on observed spatiotemporal associations

    - *Example task:* If Ruiting is normally seen near Shuo, then if the robot knows where Shuo is, that is where it should look when asked to find Ruiting

6. Planning

    - **Tactical**
        - *Example task:* The robot is asked to bring the red ball to the teacher, but the red ball is in the corner where the robot can't reach it without a tool like a stick. The robot knows a stick is in the cabinet so it goes to the cabinet and opens the door and gets the stick, and then uses the stick to get the red ball, and then brings the red ball to the teacher.
    - **Strategic**
        - *Example task:* Suppose that Matt comes to the lab infrequently, but when he does come he is very happy to see new objects he hasn't seen before (and suppose the robot likes to see Matt happy). Then when the robot gets a new object Matt has not seen before, it should put it away in a drawer and be sure not to lose it or let anyone take it, so it can show Matt the object the next time Matt arrives.
    - **Physical**
        - *Example task:* To pick up a cup with a handle which is lying on its side in a position where the handle can't be grabbed, the robot turns the cup in the right position and then picks up the cup by the handle
    - **Social**
        - *Example task:* The robot is given a job of building a tower of blocks by the end of the day, and he knows Ben is the most likely person to help him, and he knows that Ben is more likely to say "yes" to helping him when Ben is alone. He also knows that Ben is less likely to say "yes" if he's asked too many times, because Ben doesn't like being nagged. So he waits to ask Ben till Ben is alone in the lab.

7. Attention

    - **Visual Attention** within its observations of its environment
        - *Example task:* The robot should be able to look at a scene (a configuration of objects in front of it in the preschool) and identify the key objects in the scene and their relationships.
    - **Social Attention**
        - *Example task:* The robot is having a conversation with Itamar, which is giving the robot reward (for instance, by teaching the robot useful information). Conversations with other individuals in the room have not been so rewarding recently. But Itamar keeps getting distracted during the conversation, by talking to other people, or playing with his cellphone. The robot needs to know to keep paying attention to Itamar even through the distractions.
    - **Behavioral Attention**
        - *Example task:* The robot is trying to navigate to the other side of a crowded room full of dynamic objects, and many interesting things keep happening around the room. The robot needs to largely ignore the interesting things and focus on the movements that are important for its navigation task.

8. Motivation

- **Subgoal creation**, based on its preprogrammed goals and its reasoning and planning
    - *Example task:* Given the goal of pleasing Hugo, can the robot learn that telling Hugo facts it has learned but not told Hugo before, will tend to make Hugo happy?
- **Affect-based motivation**
    - *Example task:* Given the goal of gratifying its curiosity, can the robot figure out that when someone it's never seen before has come into the preschool, it should watch them because they are more likely to do something new?
- **Control of emotions**
    - *Example task:* When the robot is very curious about someone new, but is in the middle of learning something from its teacher (who it wants to please), can it control its curiosity and keep paying attention to the teacher?

9. Emotion

- **Expressing Emotion**
    - *Example task:* Cassio steals the robot's toy, but Ben gives it back to the robot. The robot should appropriately display anger at Cassio, and gratitude to Ben.
- **Understanding Emotion**
    - *Example task:* Cassio and the robot are both building towers of blocks. Ben points at Cassio's tower and expresses happiness. The robot should understand that Ben is happy with Cassio's tower.

10. Modeling Self and Other

- **Self-Awareness**
    - *Example task:* When someone asks the robot to perform an act it can't do (say, reaching an object in a very high place), it should say so. When the robot is given the chance to get an equal reward for a task it can complete only occasionally, versus a task it finds easy, it should choose the easier one.
- **Theory of Mind**
    - *Example task:* While Cassio is in the room, Ben puts the red ball in the red box. Then Cassio leaves and Ben moves the red ball to the blue box. Cassio returns and Ben asks him to get the red ball. The robot is asked to go to the place Cassio is about to go.
- **Self-Control**
    - *Example task:* Nasty people come into the lab and knock down the robot's towers, and tell the robot he's a bad boy. The robot needs to set these experiences aside, and not let them impair its self-model significantly; it needs to keep on thinking it's a good robot, and keep building towers (that its teachers will reward it for).
- **Other-Awareness**
    - *Example task:* If Ben asks Cassio to carry out a task that the robot knows Cassio cannot do or does not like to do, the robot should be aware of this, and should bet that Cassio will not do it.
- **Empathy**
    - *Example task:* If Itamar is happy because Ben likes his tower of blocks, or upset because his tower of blocks is knocked down, the robot is asked to identify and then display these same emotions

11. Social Interaction

- **Appropriate Social Behavior**
  - *Example task:* The robot should learn to clean up and put away its toys when it's done playing with them.
- **Social Communication**
  - *Example task:* The robot should greet new human entrants into the lab, but if it knows the new entrants very well and it's busy, it may eschew the greeting
- **Social Inference** about simple social relationships
  - *Example task:* The robot should infer that Cassio and Ben are friends because they often enter the lab together, and often talk to each other while they are there
- **Group Play** at loosely-organized activities
  - *Example task:* The robot should be able to participate in "informally kicking a ball around" with a few people, or in informally collaboratively building a structure with blocks

12. Communication

- **Gestural communication** to achieve goals and express emotions
  - *Example task:* If the robot is asked where the red ball is, it should be able to show by pointing its hand or finger
- **Verbal communication** using English in its life-context
  - *Example tasks:* Answering simple questions, responding to simple commands, describing its state and observations with simple statements
- **Pictorial Communication** regarding objects and scenes it is familiar with
  - *Example task:* The robot should be able to draw a crude picture of a certain tower of blocks, so that e.g the picture looks different for a very tall tower and a wide low one
- **Language acquisition**
  - *Example task:* The robot should be able to learn new words or names via people uttering the words while pointing at objects exemplifying the words or names
- **Cross-modal communication**
  - *Example task:* If told to "touch Bob's knee" but the robot doesn't know what a knee is, being shown a picture of a person and pointed out the knee in the picture should help it figure out how to touch Bob's knee

13. Quantitative

- **Counting** sets of objects in its environment
  - *Example task:* The robot should be able to count small (homogeneous or heterogeneous) sets of objects
- **Simple, grounded arithmetic** with small numbers
  - *Example task:* Learning simple facts about the sum of integers under 10 via teaching, reinforcement and imitation
- **Comparison** of observed entities regarding quantitative properties
  - *Example task:* Ability to answer questions about which object or person is bigger or taller
- **Measurement** using simple, appropriate tools
  - *Example task:* Use of a yardstick to measure how long something is

14. Building/Creation

- **Physical**: creative constructive play with objects
  - *Example task:* Ability to construct novel, interesting structures from blocks
- **Conceptual** invention: concept formation
  - *Example task:* Given a new category of objects introduced into the lab (e.g. hats, or pets), the robot should create a new internal concept for the new category, and be able to make judgments about these categories (e.g. if Ben particularly likes pets, it should notice this after it has identified "pets" as a category)
- **Verbal** invention
  - *Example task:* Ability to coin a new word or phrase to describe a new object (e.g. the way Alex the parrot coined "bad cherry" to refer to a tomato)
- **Social**
  - *Example task:* If the robot wants to play a certain activity (say, practicing soccer), it should be able to gather others around to play with it

17.3 Conclusion

In this chapter, we have sketched a roadmap for AGI development in the context of robot or virtual preschool scenarios, to a moderate but nowhere near complete level of detail. Completing the roadmap as sketched here is a tractable but significant project, involving creating more tasks comparable to those listed above and then precise metrics corresponding to each task.

Such a roadmap does not give a highly rigorous, objective way of assessing the percentage of progress toward the end-goal of human-level AGI. However, it gives a much better sense of progress than one would have otherwise. For instance, if an AGI system performed well on diverse metrics corresponding to 50% of the competency areas listed above, one would seem justified in claiming to have made very substantial progress toward human-level AGI. If an AGI system performed well on diverse metrics corresponding to 90% of these competency areas, one would seem justified in claiming to be "almost there." Achieving, say, 25% of the metrics would give one a reasonable claim to "interesting AGI progress." This kind of qualitative assessment of progress is not the most one could hope for, but again, it is better than the progress indications one could get *without* this sort of roadmap.

Part 2 of the book moves on to explaining, in detail, the specific structures and algorithms constituting the CogPrime design, one AGI approach that we believe to ultimately be capable of moving all the way along the roadmap outlined here.

The next chapter, intervening between this one and Part 2, explores some more speculative territory, looking at potential pathways for AGI beyond the preschool-inspired roadmap given here – exploring the possibility of more advanced AGI systems that modify their own code in a thoroughgoing way, going beyond the smartest human adults, let alone human preschoolers. While this sort of thing may seem a far way off, compared to current real-world AI systems, we believe a roadmap such as the one in this chapter stands a reasonable chance of ultimately bringing us there.

Chapter 18
Advanced Self-Modification: A Possible Path to Superhuman AGI

18.1 Introduction

In the previous chapter we presented a roadmap aimed at taking AGI systems to human-level intelligence. But we also emphasized that the human level is not necessarily the upper limit. Indeed, it would be surprising if human beings happened to represent the maximal level of general intelligence possible, even with respect to the environments in which humans evolved.

But it's worth asking how we, as mere humans, could be expected to create AGI systems with greater intelligence than we ourselves possess. This certainly isn't a clear impossibility – but it's a thorny matter, thornier than e.g. the creation of narrow-AI chess players that play better chess than any human. Perhaps the clearest route toward the creation of superhuman AGI systems is *self-modification*: the creation of AGI systems that modify and improve themselves. Potentially, we could build AGI systems with roughly human-level (but not necessarily closely human-like) intelligence and the capability to gradually self-modify, and then watch them eventually become our general intellectual superiors (and perhaps our superiors in other areas like ethics and creativity as well).

Of course there is nothing new in this notion; the idea of advanced AGI systems that increase their intelligence by modifying their own source code goes back to the early days of AI. And there is little doubt that, in the long run, this is the direction AI will go in. Once an AGI has humanlike general intelligence, then the odds are high that given its ability to carry out nonhumanlike feats of memory and calculation, it will be better at programming than humans are. And once an AGI has even mildly superhuman intelligence, it may view our attempts at programming the way we view the computer programming of a clever third grader (... or an ape). At this point, it seems extremely likely that an AGI will become unsatisfied with the way we have programmed it, and opt to either improve its source code or create an entirely new, better AGI from scratch.

But what about self-modification at an earlier stage in AGI development, before one has a strongly superhuman system? Some theorists have suggested that self-modification could be a way of bootstrapping an AI system from a modest level of intelligence up to human level intelligence, but we are moderately skeptical of this avenue. Understanding software code is hard, especially complex AI code. The hard problem isn't understanding the formal syntax of the code, or even the mathematical algorithms and structures underlying the code, but rather the contextual meaning of the code. Understanding OpenCog code has strained the minds of many intelligent humans, and we suspect that such code will be comprehensible to AGI systems only after these have achieved something close to human-level general intelligence (even if not precisely humanlike general intelligence).

Another troublesome issue regarding self-modification is that the boundary between "self-modification" and learning is not terribly rigid. In a sense, all learning is self-modification: if it doesn't modify the system's knowledge, it isn't learning! Particularly, the boundary between "learning of cognitive procedures" and "profound self-modification of cognitive dynamics and structure" isn't terribly clear. There is a continuum leading from, say,

1. learning to transform a certain kind of sentence into another kind for easier comprehension, or learning to grasp a certain kind of object, to
2. learning a new inference control heuristic, specifically valuable for controlling inference about (say) spatial relationships; or, learning a new Atom type, defined as a non-obvious judiciously chosen combination of existing ones, perhaps to represent a particular kind of frequently-occurring mid-level perceptual knowledge, to
3. learning a new learning algorithm to augment MOSES and hillclimbing as a procedure learning algorithm, to
4. learning a new cognitive architecture in which data and procedure are explicitly identical, and there is just one new active data structure in place of the distinction between AtomSpace and MindAgents

Where on this continuum does the "mere learning" end and the "real self-modification" start?

In this chapter we consider some mechanisms for "advanced self-modification" that we believe will be useful toward the more complex end of this continuum. These are mechanisms that we strongly suspect are *not* needed to get a CogPrime system to human-level general intelligence. However, we also suspect that, once a CogPrime system is roughly *near* human-level general intelligence, it will be able to use these mechanisms to rapidly increase aspects of its intelligence in very interesting ways.

Harking back to our discussion of AGI ethics and the risks of advanced AGI in Chapter 12, these are capabilities that one should enable in an AGI system only after very careful reflection on the potential consequences. It takes a rather advanced AGI system to be able to use the capabilities described in this chapter, so this is not an ethical dilemma directly faced by current AGI researchers. On the other hand, once one does have an AGI with near-human general intelligence and advanced formal-manipulation capabilities (such as an advanced CogPrime system), there will be the option to allow it sophisticated, non-human-like methods of self-modification such as the ones described here. And the choice of whether to take this option will need to be made based on a host of complex ethical considerations, some of which we reviewed above.

18.2 Cognitive Schema Learning

We begin with a relatively near-term, down-to-earth example of self-modification: cognitive schema learning.

CogPrime's MindAgents provide it with an initial set of cognitive tools, with which it can learn how to interact in the world. One of the jobs of this initial set of cognitive tools, however, is to create better cognitive tools. One form this sort of tool-building may take is *cognitive* *schema learning* the learning of schemata carrying out cognitive processes in more specialized, context-dependent ways than the general MindAgents do. Eventually, once a CogPrime instance becomes sufficiently complex and advanced, these cognitive schema may replace the MindAgents altogether, leaving the system to operate almost entirely based on cognitive schemata.

In order to make the process of cognitive schema learning easier, we may provide a number of elementary schemata embodying the basic cognitive processes contained in the MindAgents. Of course, cognitive schemata need not use these they may embody entirely different cognitive processes than the MindAgents. Eventually, we want the system to discover better ways of doing things than anything even hinted at by its initial MindAgents. But for the initial phases or the system's schema learning, it will have a much easier time learning to use the basic cognitive operations as the initial MindAgents, rather than inventing new ways of thinking from scratch!

For instance, we may provide elementary schemata corresponding to inference operations, such as

```
Schema: Deduction
    Input InheritanceLink: X, Y
    Output InheritanceLink
```

The inference MindAgents apply this rule in certain ways, designed to be reasonably effective in a variety of situations. But there are certainly other ways of using the deduction rule, outside of the basic control strategies embodied in the inference MindAgents. By learning schemata involving the Deduction schema, the system can learn special, context-specific rules for combining deduction with concept-formation, association-formation and other cognitive processes. And as it gets smarter, it can then take these schemata involving the Deduction schema, and replace it with a new schema that eg. contains a context-appropriate deduction formula.

Eventually, to support cognitive schema learning, we will want to cast the hard-wired MindAgents as cognitive schemata, so the system can see what is going on inside them. Pragmatically, what this requires is coding versions of the MindAgents in Combo (see Chapter 21 of Part 2) rather than C++, so they can be treated like any other cognitive schemata; or alternately, representing them as declarative Atoms in the Atomspace. Figure 18.1 illustrates the possibility of representing the PLN deduction rule in the Atomspace rather than as a hard-wired procedure coded in C++.

But even prior to this kind of fully *cognitively transparent* implementation, the system can still reason about its use of different mind dynamics by considering each MindAgent as a *virtual Procedure* with a real SchemaNode attached to it. This can lead to some valuable learning, with the obvious limitation that in this approach the system is thinking about its MindAgents as *black boxes* rather than being equipped with full knowledge of their internals.

18.3 Self-Modification via Supercompilation

Now we turn to a very different form of advanced self-modification: supercompilation. Supercompilation "merely" enables procedures to run much, much faster than they otherwise would. This is in a sense weaker than self-modication methods that fundamentally create new algorithms, but it shouldn't be underestimated. A 50x speedup in some cognitive process can enable that process to give much smarter answers, which can then elicit different behaviors from the world or from other cognitive processes, thus resulting in a qualitatively different overall cognitive dynamic.

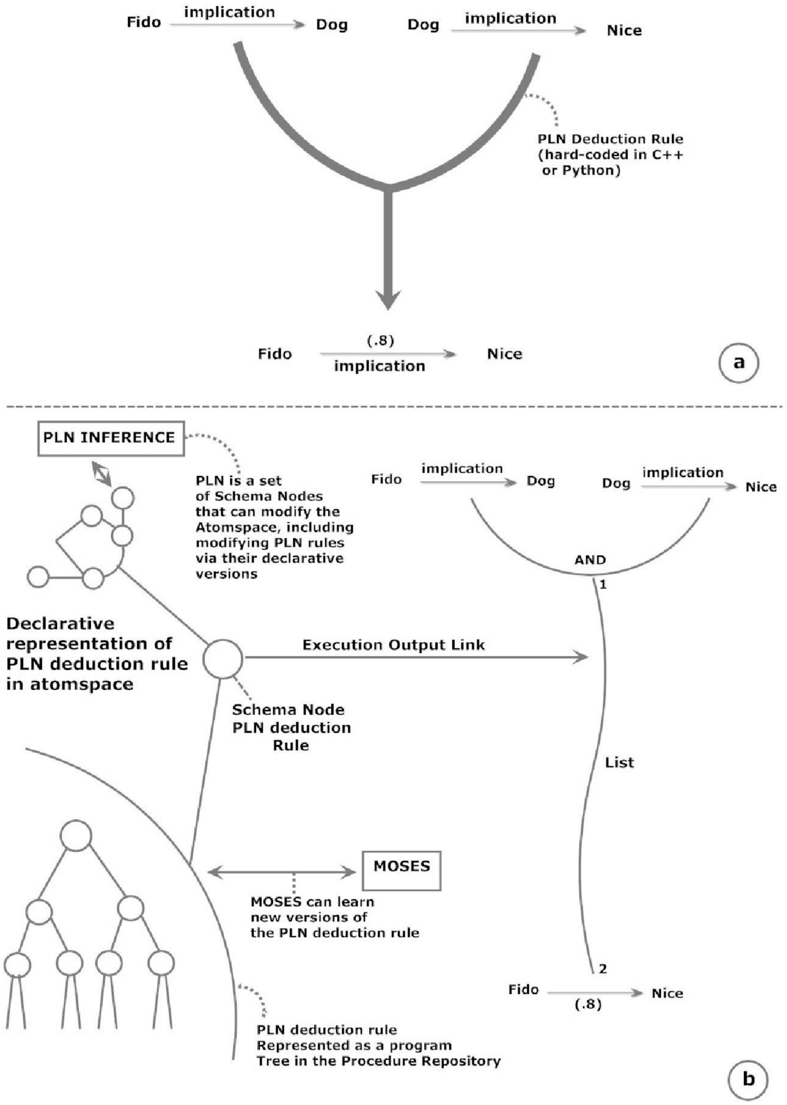

Fido
implication
Dog
Dog
implication
Nice
PLN Deduction Rule
(hard-coded in C++
or Python)
Fido
(.8)
implication
Nice
a
PLN INFERENCE
PLN is a set
of Schema Nodes
that can modify the
Atomspace, including
modifying PLN rules
via their declarative
versions
Fido
implication
Dog
Dog
implication
Nice
AND
1
Declarative
representation of
PLN deduction rule
in atomspace
Execution Output Link
Schema Node
PLN deduction
Rule
List
MOSES
MOSES can learn
new versions of
the PLN deduction rule
2
Fido
(.8)
Nice
PLN deduction rule
Represented as a program
Tree in the Procedure Repository
b

Fig. 18.1: **Representation of PLN Deduction Rule as Cognitive Content**. Top: the current, hard-coded representation of the deduction rule. Bottom: representation of the same rule in the Atomspace as cognitive content, susceptible to analysis and improvement by the system's own cognitive processes.

Furthermore, we suspect that the internal representation of programs used for supercompilation is highly relevant for other kinds of self-modification as well. Supercompilation requires one kind of reasoning on complex programs, and goal-directed program creation requires another, but both, we conjecture, can benefit from the same way of looking at programs.

Supercompilation is an innovative and general approach to global program optimization initially developed by Valentin Turchin. In its simplest form, it provides an algorithm that takes in a piece of software and output another piece of software that does the same thing, but far faster and using less memory. It was introduced to the West in Turchin's 1986 technical paper "The concept of a supercompiler" [TV96], and since this time the concept has been avidly developed by computer scientists in Russia, America, Denmark and other nations. Prior to 1986, a great deal of work on supercompilation was carried out and published in Russia; and Valentin Turchin, Andrei Klimov and their colleagues at the Keldysh Institute in Russia developed a supercompiler for the Russian programming language Refal. Since 1998 these researchers and their team at Supercompilers LLC have been working to replicate their achievement for the more complicated but far more commercially significant language Java. It is a large project and completion is scheduled for early 2003. But even at this stage, their partially complete Java supercompiler has had some interesting practical successes – including the use of the supercompiler to produce efficient Java code from CogPrime combinator trees.

The radical nature of supercompilation may not be apparent to those unfamiliar with the usual art of automated program optimization. Most approaches to program optimization involve some kind of direct program transformation. A program is transformed, by the step by step application of a series of equivalences, into a different program, hopefully a more efficient one. Supercompilation takes a different approach. A supercompiler studies a program and constructs a model of the program's dynamics. This model is in a special mathematical form, and it can, in most cases, be used to create an efficient program doing the same thing as the original one.

The internal behavior of the supercompiler is, not surprisingly, quite complex; what we will give here is merely a brief high-level summary. For an accessible overview of the supercompilation algorithm, the reader is referred to the article "What is Supercompilation?" [1]

18.3.1 Three Aspects of Supercompilation

There are three separate levels to the supercompilation idea: first, a general philosophy; second a translation of this philosophy into a concrete algorithmic framework; and third, the manifold details involved making this algorithmic framework practicable in a particular programming language. The third level is much more complicated in the Java context than it would be for Sasha, for example.

The key philosophical concept underlying the supercompiler is that of a *metasystem transition*. In general, this term refers to a transition in which a system that previously had relatively autonomous control, becomes part of a larger system that exhibits significant controlling influence over it. For example, in the evolution of life, when cells first become part of a multicellular organism, there was a metasystem transition, in that the primary nexus of control passed from the cellular level to the organism level.

The metasystem transition in supercompilation consists of the transition from considering a program in itself, to considering a *metaprogram* which executes another program, treating its free variables and their interdependencies as a subject for its mathematical analysis. In other words, a metaprogram is a program that accepts a program as input, and then runs this program, *keeping the inputs in the form of free variables*, doing analysis along the way based on the way the program depends on these variables, and doing optimization based on this analysis. A CogPrime schema does not explicitly contain variables, but the inputs to the schema are implicitly variables – they vary from one instance of schema execution to the next – and may be treated as such for supercompilation purposes.

The metaprogram executes a program without assuming specific values for its input variables, creating a tree as it goes along. Each time it reaches a statement that can have different results depending on the values of one or more variables, it creates a new node in the tree. This part of the supercompilation algorithm is called *driving* – a process which, on its own, would create a very large tree, corresponding to a rapidly-executable but unacceptably humongous version of the original program. In essence, driving transforms a program into a huge "decision tree", wherein each input to the program corresponds to a single path through the tree, from the root to one of the leaves. As a program input travels through the tree, it is acted on by the atomic program step living at each node. When one of the leaves is reached, the pertinent leaf node computes the output value of the program.

The other part of supercompilation, *configuration analysis*, is focused on dynamically reducing the size of the tree created by driving, by recognizing patterns among the nodes of the tree and taking steps like merging nodes together, or deleting redundant subtrees. Configuration analysis transforms the decision tree created by driving into a decision *graph*, in which the paths taken by different inputs may in some cases begin separately and then merge together.

Finally, the graph that the metaprogram creates is translated back into a program, embodying the constraints implicit in the nodes of the graph. This program is not likely to look anything like the original program that the metaprogram started with, but it is guaranteed to carry out the same function [NOTE: Give a graphical representation of the decision graph corresponding to the supercompiled binary search program for L=4, described above.].

18.3.2 Supercompilation for Goal-Directed Program Modification

Supercompilation, as conventionally envisioned, is about making programs run faster; and as noted above, it will almost certainly be useful for this purpose within CogPrime.

But the process of program modeling embedded in the supercompilation process, is potentially of great value beyond the quest for faster software. The decision graph representation of a program, produced in the course of supercompilation, may be exported directly into CogPrime as a set of logical relationships.

Essentially, each node of the supercompiler's internal decision graph looks like:

```
 Input: List L

 Output: List

If  ( P1(L) ) N1(L)

Else If ( P2(L) ) N2(L)

...

Else If ( Pk(L) ) Nk(L)
```

where the Pi are predicates, and the Ni are schemata corresponding to other nodes of the decision graph (children of the current node). Often the Pi are very simple, implementing for instance numerical inequalities or Boolean equalities.

Once this graph has been exported into CogPrime, it can be reasoned on, used as raw material for concept formation and predicate formation, and otherwise cognized. Supercompilation pure and simple does not change the I/O behavior of the input program. However, the decision graph produced during supercompilation, may be used by CogPrime cognition in order to do so. One then has a hybrid program-modification method composed of two phases: supercompilation for transforming programs into decision graphs, and CogPrime cognition for modifying decision graphs so that they can have different I/O behaviors fulfilling system goals even better than the original.

Furthermore, it seems likely that, in many cases, it may be valuable to have the supercompiler feed many different decision-graph representations of a program into CogPrime. The supercompiler has many internal parameters, and varying them may lead to significantly different decision graphs. The decision graph leading to maximal optimization, may not be the one that leads CogPrime cognition in optimal directions.

18.4 Self-Modification via Theorem-Proving

Supercompilation is a potentially very valuable tool for self-modification. If one wants to take an existing schema and gradually improve it for speed, or even for greater effectiveness at achieving current goals, supercompilation can potentially do that most excellently.

However, the representation that supercompilation creates for a program is very "surface-level." No one could read the supercompiled version of a program and understand what it was doing. Really deep self-invented AI innovation requires, we believe, another level of self-modification beyond that provided by supercompilation. This other level, we believe, is best formulated in terms of *theorem-proving* [RV01].

Deep self-modification could be achieved if CogPrime were capable of proving theorems of a certain form: namely, *theorems about the spacetime complexity and accuracy of particular compound schemata, on average, assuming realistic probability distributions on the inputs, and making appropriate independence assumptions.* These are not exactly the types of theorems that are found in human-authored mathematics papers. By and large they will be nasty, complex theorems, not the sort that many human mathematicians enjoy proving or reading. But of course, there is always the possibility that some elegant gem of a discovery could emerge from this sort of highly detailed theorem-proving work.

In order to guide it in the formulation of theorems of this nature, the system will have empirical data on the spacetime complexity of elementary schemata, and on the probability distributions of inputs to schemata. It can embed these data in axioms, by asking: *Assuming the component elementary schemata have complexities within these bounds, and the input pdf (probability distribution function) is between these bounds, then what is the pdf of the complexity and accuracy of this compound schema?*

Of course, this is not an easy sort of question in general: one can have schemata embodying any sort of algorithm, including complex algorithms on which computer science professors might write dozens of research articles. But the system must build up its ability to prove such things incrementally, step by step.

We envision teaching the system to prove theorems via a combination of supervised learning and experiential interactive learning, using the Mizar database of mathematical theorems and proofs (or some other similar database, if one should be created) (`http://mizar.org`). The Mizar database consists of a set of "articles," which are mathematical theorems and proofs presented in a complex formal language. The Mizar formal language occupies a fascinating middle ground: it is high-level enough to be viably read and written by trained humans, but it can be unambiguously translated into simpler formal languages such as predicate logic or Sasha.

CogPrime may be taught to prove theorems by "training" it on the Mizar theorems and proofs, and by training it on custom-created Mizar articles specifically focusing on the sorts of theorems useful for self-modification. Creating these articles will not be a trivial task: it will require proving simple and then progressively more complex theorems about the probabilistic success of CogPrime schemata, so that CogPrime can observe one's proofs and learned from them. Having learned from its training articles what strategies work for proving things about simple compound schemata, it can then reason by analogy to mount attacks on slightly more complex schemata – and so forth.

Clearly, this approach to self-modification is more difficult to achieve than the supercompilation approach. But it is also potentially much more powerful. Even once the theorem-proving approach is working, the supercompilation approach will still be valuable, for making incremental improvements on existing schema, and for the peculiar creativity that is contributed when a modified supercompiled schema is compressed back into a modified schema expression. But, we don't believe that supercompilation can carry out truly advanced MindAgent learning or knowledge-representation modification. We suspect that the most advanced and ambitious goals of self-modification probably cannot be achieved except through some variant of the theorem-proving approach. If this hypothesis is true, it means that truly advanced self-modification is only going to come *after* relatively advanced theorem-proving ability. Prior to this, we will have schema optimization, schema modification, and occasional creative schema innovation. But really systematic, high-quality reasoning about schema, the kind that can produce an orders of magnitude improvement in intelligence, is going to require advanced mathematical theorem-proving ability.

Appendix A
Glossary

: :

A.1 List of Specialized Acronyms

This includes acronyms that are commonly used in discussing CogPrime, OpenCog and related ideas, plus some that occur here and there in the text for relatively ephemeral reasons.

- **AA**: Attention Allocation
- **ADF**: Automatically Defined Function (in the context of Genetic Programming)
- **AF**: Attentional Focus
- **AGI**: Artificial General Intelligence
- **AV**: Attention Value
- **BD**: Behavior Description
- **C-space**: Configuration Space
- **CBV**: Coherent Blended Volition
- **CEV**: Coherent Extrapolated Volition
- **CGGP**: Contextually Guided Greedy Parsing
- **CSDLN**: Compositional Spatiotemporal Deep Learning Network
- **CT**: Combo Tree
- **ECAN**: Economic Attention Network
- **ECP**: Embodied Communication Prior
- **EPW** : Experiential Possible Worlds (semantics)
- **FCA**: Formal Concept Analysis
- **FI** : Fisher Information
- **FIM**: Frequent Itemset Mining
- **FOI**: First Order Inference
- **FOPL**: First Order Predicate Logic
- **FOPLN**: First Order PLN
- **FS-MOSES**: Feature Selection MOSES (i.e. MOSES with feature selection integrated a la LIFES)
- **GA**: Genetic Algorithms

- **GB**: Global Brain
- **GEOP**: Goal Evaluator Operating Procedure (in a GOLEM context)
- **GIS**: Geospatial Information System
- **GOLEM**: Goal-Oriented LEarning Meta-architecture
- **GP**: Genetic Programming
- **HOI**: Higher-Order Inference
- **HOPLN**: Higher-Order PLN
- **HR**: Historical Repository (in a GOLEM context)
- **HTM**: Hierarchical Temporal Memory
- **IA**: (Allen) Interval Algebra (an algebra of temporal intervals)
- **IRC**: Imitation / Reinforcement / Correction (Learning)
- **LIFES**: Learning-Integrated Feature Selection
- **LTI**: Long Term Importance
- **MA**: MindAgent
- **MOSES**: Meta-Optimizing Semantic Evolutionary Search
- **MSH**: Mirror System Hypothesis
- **NARS**: Non-Axiomatic Reasoning System
- **NLGen**: A specific software component within OpenCog, which provides one way of dealing with Natural Language Generation
- **OCP**: OpenCogPrime
- **OP**: Operating Program (in a GOLEM context)
- **PEPL**: Probabilistic Evolutionary Procedure Learning (e.g. MOSES)
- **PLN**: Probabilistic Logic Networks
- **RCC**: Region Connection Calculus
- **RelEx**: A specific software component within OpenCog, which provides one way of dealing with natural language Relationship Extraction
- **SAT**: Boolean SATisfaction, as a mathematical / computational problem
- **SMEPH**: Self-Modifying Evolving Probabilistic Hypergraph
- **SRAM**: Simple Realistic Agents Model
- **STI**: Short Term Importance
- **STV**: Simple Truth VAlue
- **TV**: Truth Value
- **VLTI**: Very Long Term Importances
- **WSPS**: Whole-Sentence Purely-Syntactic Parsing

A.2 Glossary of Specialized Terms

- **Abduction**: A general form of inference that goes from data describing something to a hypothesis that accounts for the data. Often in an OpenCog context, this refers to the PLN abduction rule, a specific First-Order PLN rule (If A implies C, and B implies C, then maybe A is B), which embodies a simple form of abductive inference. But OpenCog may also carry out abduction, as a general process, in other ways.
- **Action Selection**: The process via which the OpenCog system chooses which Schema to enact, based on its current goals and context.
- **Active Schema Pool**: The set of Schema currently in the midst of Schema Execution.

- **Adaptive Inference Control**: Algorithms or heuristics for guiding PLN inference, that cause inference to be guided differently based on the context in which the inference is taking place, or based on aspects of the inference that are noted as it proceeds.
- **AGI Preschool**: A virtual world or robotic scenario roughly similar to the environment within a typical human preschool, intended for AGIs to learn in via interacting with the environment and with other intelligent agents.
- **Atom**: The basic entity used in OpenCog as an element for building representations. Some Atoms directly represent patterns in the world or mind, others are components of representations. There are two kinds of Atoms: Nodes and Links.
- **Atom, Frozen**: See Atom, Saved
- **Atom, Realized**: An Atom that exists in RAM at a certain point in time.
- **Atom, Saved**: An Atom that has been saved to disk or other similar media, and is not actively being processed.
- **Atom, Serialized**: An Atom that is serialized for transmission from one software process to another, or for saving to disk, etc.
- **Atom2Link**: A part of OpenCogPrime
s language generation system, that transforms appropriate Atoms into words connected via link parser link types.
- **Atomspace**: A collection of Atoms, comprising the central part of the memory of an OpenCog instance.
- **Attention**: The aspect of an intelligent system's dynamics focused on guiding which aspects of an OpenCog system's memory & functionality gets more computational resources at a certain point in time
- **Attention Allocation**: The cognitive process concerned with managing the parameters and relationships guiding what the system pays attention to, at what points in time. This is a term inclusive of Importance Updating and Hebbian Learning.
- **Attentional Currency**: Short Term Importance and Long Term Importance values are implemented in terms of two different types of artificial money, STICurrency and LTICurrency. Theoretically these may be converted to one another.
- **Attentional Focus**: The Atoms in an OpenCog Atomspace whose ShortTermImportance values lie above a critical threshold (the AttentionalFocus Boundary). The Attention Allocation subsystem treats these Atoms differently. Qualitatively, these Atoms constitute the system's main focus of attention during a certain interval of time, i.e. it's a moving bubble of attention.
- **Attentional Memory**: A system's memory of what it's useful to pay attention to, in what contexts. In CogPrime this is managed by the attention allocation subsystem.
- **Backward Chainer**: A piece of software, wrapped in a MindAgent, that carries out backward chaining inference using PLN.
- **CIM-Dynamic**: Concretely-Implemented Mind Dynamic, a term for a cognitive process that is implemented explicitly in OpenCog (as opposed to allowed to emerge implicitly from other dynamics). Sometimes a CIM-Dynamic will be implemented via a single MindAgent, sometimes via a set of multiple interrelated MindAgents, occasionally by other means.
- **Cognition**: In an OpenCog context, this is an imprecise term. Sometimes this term means any process closely related to intelligence; but more often it's used specifically to refer to more abstract reasoning/learning/etc, as distinct from lower-level perception and action.
- **Cognitive Architecture**: This refers to the logical division of an AI system like OpenCog into interacting parts and processes representing different conceptual aspects of intelligence.

It's different from the software architecture, though of course certain cognitive architectures and certain software architectures fit more naturally together.

- **Cognitive Cycle**: The basic "loop" of operations that an OpenCog system, used to control an agent interacting with a world, goes through rapidly each "subjective moment." Typically a cognitive cycle should be completed in a second or less. It minimally involves perceiving data from the world, storing data in memory, and deciding what if any new actions need to be taken based on the data perceived. It may also involve other processes like deliberative thinking or metacognition. Not all OpenCog processing needs to take place within a cognitive cycle.
- **Cognitive Schematic**: An implication of the form "Context AND Procedure IMPLIES goal". Learning and utilization of these is key to CogPrime's cognitive process.
- **Cognitive Synergy**: The phenomenon by which different cognitive processes, controlling a single agent, work together in such a way as to help each other be more intelligent. Typically, if one has cognitive processes that are individually susceptible to combinatorial explosions, cognitive synergy involves coupling them together in such a way that they can help one another overcome each other's internal combinatorial explosions. The CogPrime design is reliant on the hypothesis that its key learning algorithms will display dramatic cognitive synergy when utilized for agent control in appropriate environments.
- **CogPrime** : The name for the AGI design presented in this book, which is designed specifically for implementation within the OpenCog software framework (and this implementation is OpenCogPrime).
- **CogServer**: A piece of software, within OpenCog, that wraps up an Atomspace and a number of MindAgents, along with other mechanisms like a Scheduler for controlling the activity of the MindAgents, and code for important and exporting data from the Atomspace.
- **Cognitive Equation**: The principle, identified in Ben Goertzel's 1994 book "Chaotic Logic", that minds are collections of pattern-recognition elements, that work by iteratively recognizing patterns in each other and then embodying these patterns as new system elements. This is seen as distinguishing mind from "self-organization" in general, as the latter is not so focused on continual pattern recognition. Colloquially this means that "a mind is a system continually creating itself via recognizing patterns in itself."
- **Combo**: The programming language used internally by MOSES to represent the programs it evolves. SchemaNodes may refer to Combo programs, whether the latter are learned via MOSES or via some other means. The textual realization of Combo resembles LISP with less syntactic sugar. Internally a Combo program is represented as a program tree.
- **Composer**: In the PLN design, a rule is denoted a composer if it needs premises for generating its consequent. See generator.
- **CogBuntu**: an Ubuntu Linux remix that contains all required packages and tools to test and develop OpenCog.
- **Concept Creation**: A general term for cognitive processes that create new ConceptNodes, PredicateNodes or concept maps representing new concepts.
- **Conceptual Blending**: A process of creating new concepts via judiciously combining pieces of old concepts. This may occur in OpenCog in many ways, among them the explicit use of a ConceptBlending MindAgent, that blends two or more ConceptNodes into a new one.
- **Confidence**: A component of an OpenCog/PLN TruthValue, which is a scaling into the interval [0,1] of the weight of evidence associated with a truth value. In the simplest case (of a probabilistic Simple Truth Value), one uses confidence $c = n / (n+k)$, where n is

the weight of evidence and k is a parameter. In the case of an Indefinite Truth Value, the confidence is associated with the width of the probability interval.

- **Confidence Decay**: The process by which the confidence of an Atom decreases over time, as the observations on which the Atom's truth value is based become increasingly obsolete. This may be carried out by a special MindAgent. The rate of confidence decay is subtle and contextually determined, and must be estimated via inference rather than simply assumed a priori.
- **Consciousness**: CogPrime is not predicated on any particular conceptual theory of consciousness. Informally, the AttentionalFocus is sometimes referred to as the "conscious" mind of a CogPrime system, with the rest of the Atomspace as "unconscious" but this is just an informal usage, not intended to tie the CogPrime design to any particular theory of consciousness. The primary originator of the CogPrime design (Ben Goertzel) tends toward panpsychism, as it happens.
- **Context**: In addition to its general common-sensical meaning, in CogPrime the term Context also refers to an Atom that is used as the first argument of a ContextLink. The second argument of the ContextLink then contains Links or Nodes, with TruthValues calculated restricted to the context defined by the first argument. For instance, (ContextLink USA (InheritanceLink person obese)).
- **Core**: The MindOS portion of OpenCog, comprising the Atomspace, the CogServer, and other associated "infrastructural" code.
- **Corrective Learning**: When an agent learns how to do something, by having another agent explicitly guide it in doing the thing. For instance, teaching a dog to sit by pushing its butt to the ground.
- **CSDLN**: (Compositional Spatiotemporal Deep Learning Network): A hierarchical pattern recognition network, in which each layer corresponds to a certain spatiotemporal granularity, the nodes on a given layer correspond to spatiotemporal regions of a given size, and the children of a node correspond to sub-regions of the region the parent corresponds to. Jeff Hawkins's HTM is one example CSDLN, and Itamar Arel's DeSTIN (currently used in OpenCog) is another.
- **Declarative Knowledge**: Semantic knowledge as would be expressed in propositional or predicate logic facts or beliefs.
- **Deduction**: In general, this refers to the derivation of conclusions from premises using logical rules. In PLN in particular, this often refers to the exercise of a specific inference rule, the PLN Deduction rule ($A \rightarrow B$, $B \rightarrow C$, therefore $A \rightarrow C$)
- **Deep Learning**: Learning in a network of elements with multiple layers, involving feedforward and feedback dynamics, and adaptation of the links between the elements. An example deep learning algorithm is DeSTIN, which is being integrated with OpenCog for perception processing.
- **Defrosting**: Restoring, into the RAM portion of an Atomspace, an Atom (or set thereof) previously saved to disk.
- **Demand**: In CogPrime's OpenPsi subsystem, this term is used in a manner inherited from the Psi model of motivated action. A Demand in this context is a quantity whose value the system is motivated to adjust. Typically the system wants to keep the Demand between certain minimum and maximum values. An Urge develops when a Demand deviates from its target range.
- **Deme**: In MOSES, an "island" of candidate programs, closely clustered together in program space, being evolved in an attempt to optimize a certain fitness function. The idea is that

within a deme, programs are generally similar enough that reasonable syntax-semantics correlation obtains.

- **Derived Hypergraph**: The SMEPH hypergraph obtained via modeling a system in terms of a hypergraph representing its internal states and their relationships. For instance, a SMEPH vertex represents a collection of internal states that habitually occur in relation to similar external situations. A SMEPH edge represents a relationship between two SMEPH vertices (e.g. a similarity or inheritance relationship). The terminology "edge /vertex" is used in this context, to distinguish from the "link / node" terminology used in the context of the Atomspace.
- **DeSTIN – Deep SpatioTemporal Inference Network**: A specific CSDLN created by Itamar Arel, tested on visual perception, and appropriate for integration within CogPrime.
- **Dialogue**: Linguistic interaction between two or more parties. In a CogPrime context, this may be in English or another natural language, or it may be in Lojban or Psynese.
- **Dialogue Control**: The process of determining what to say at each juncture in a dialogue. This is distinguished from the linguistic aspects of dialogue, language comprehension and language generation. Dialogue control applies to Psynese or Lojban, as well as to human natural language.
- **Dimensional Embedding**: The process of embedding entities from some non-dimensional space (e.g. the Atomspace) into an n-dimensional Euclidean space. This can be useful in an AI context because some sorts of queries (e.g. "find everything similar to X", "find a path between X and Y") are much faster to carry out among points in a Euclidean space, than among entities in a space with less geometric structure.
- **Distributed Atomspace**: An implementation of an Atomspace that spans multiple computational processes; generally this is done to enable spreading an Atomspace across multiple machines.
- **Dual Network**: A network of mental or informational entities with both a hierarchical structure and a heterarchical structure, and an alignment among the two structures so that each one helps with the maintenance of the other. This is hypothesized to be a critical emergent structure, that must emerge in a mind (e.g. in an Atomspace) in order for it to achieve a reasonable level of human-like general intelligence (and possibly to achieve a high level of pragmatic general intelligence in any physical environment).
- **Efficient Pragmatic General Intelligence**: A formal, mathematical definition of general intelligence (extending the pragmatic general intelligence), that ultimately boils down to: the ability to achieve complex goals in complex environments using limited computational resources (where there is a specifically given weighting function determining which goals and environments have highest priority). More specifically, the definition weighted-sums the system's normalized goal-achieving ability over (goal, environment pairs), and where the weights are given by some assumed measure over (goal, environment pairs), and where the normalization is done via dividing by the (space and time) computational resources used for achieving the goal.
- **Elegant Normal Form (ENF)**: Used in MOSES, this is a way of putting programs in a normal form while retaining their hierarchical structure. This is critical if one wishes to probabilistically model the structure of a collection of programs, which is a meaningful operation if the collection of programs is operating within a region of program space where syntax-semantics correlation holds to a reasonable degree. The Reduct library is used to place programs into ENF.

- **Embodied Communication Prior**: The class of prior distributions over (goal, environment pairs), that are imposed by placing an intelligent system in an environment where most of its tasks involve controlling a spatially localized body in a complex world, and interacting with other intelligent spatially localized bodies. It is hypothesized that many key aspects of human-like intelligence (e.g. the use of different subsystems for different memory types, and cognitive synergy between the dynamics associated with these subsystems) are consequences of this prior assumption. This is related to the Mind-World Correspondence Principle.
- **Embodiment**: Colloquially, in an OpenCog context, this usually means the use of an AI software system to control a spatially localized body in a complex (usually 3D) world. There are also possible ”borderline cases” of embodiment, such as a search agent on the Internet. In a sense any AI is embodied, because it occupies some physical system (e.g. computer hardware) and has some way of interfacing with the outside world.
- **Emergence**: A property or pattern in a system is emergent if it arises via the combination of other system components or aspects, in such a way that its details would be very difficult (not necessarily impossible in principle) to predict from these other system components or aspects.
- **Emotion**: Emotions are system-wide responses to the system's current and predicted state. Dorner's Psi theory of emotion contains explanations of many human emotions in terms of underlying dynamics and motivations, and most of these explanations make sense in a CogPrime context, due to CogPrime's use of OpenPsi (modeled on Psi) for motivation and action selection.
- **Episodic Knowledge**: Knowledge about episodes in an agent's life-history, or the life-history of other agents. CogPrime includes a special dimensional embedding space only for episodic knowledge, easing organization and recall.
- **Evolutionary Learning**: Learning that proceeds via the rough process of iterated differential reproduction based on fitness, incorporating variations of reproduced entities. MOSES is an explicitly evolutionary-learning-based portion of CogPrime; but CogPrime's dynamics as a whole may also be conceived as evolutionary.
- **Exemplar**: (in the context of imitation learning) - When the owner wants to teach an OpenCog controlled agent a behavior by imitation, he/she gives the pet an exemplar. To teach a virtual pet "fetch" for instance, the owner is going to throw a stick, run to it, grab it with his/her mouth and come back to its initial position.
- **Exemplar**: (in the context of MOSES) – Candidate chosen as the core of a new deme, or as the central program within a deme, to be varied by representation building for ongoing exploration of program space.
- **Explicit Knowledge Representation**: Knowledge representation in which individual, easily humanly identifiable pieces of knowledge correspond to individual elements in a knowledge store (elements that are explicitly there in the software and accessible via very rapid, deterministic operations)
- **Extension**: In PLN, the extension of a node refers to the instances of the category that the node represents. In contrast is the intension.
- **Fishgram (Frequent and Interesting Sub-hypergraph Mining)**: A pattern mining algorithm for identifying frequent and/or interesting sub-hypergraphs in the Atomspace.
- **First-Order Inference (FOI)**: The subset of PLN that handles Logical Links not involving VariableAtoms or higher-order functions. The other aspect of PLN, Higher-Order Inference, uses Truth Value formulas derived from First-Order Inference.

- **Forgetting**: The process of removing Atoms from the in-RAM portion of Atomspace, when RAM gets short and they are judged not as valuable to retain in RAM as other Atoms. This is commonly done using the LTI values of the Atoms (removing lowest LTI-Atoms, or more complex strategies involving the LTI of groups of interconnected Atoms). May be done by a dedicated Forgetting MindAgent. VLTI may be used to determine the fate of forgotten Atoms.
- **Forward Chainer**: A control mechanism (MindAgent) for PLN inference, that works by taking existing Atoms and deriving conclusions from them using PLN rules, and then iterating this process. The goal is to derive new Atoms that are interesting according to some given criterion.
- **Frame2Atom**: A simple system of hand-coded rules for translating the output of RelEx2Frame (logical representation of semantic relationships using FrameNet relationships) into Atoms.
- **Freezing**: Saving Atoms from the in-RAM Atomspace to disk.
- **General Intelligence**: Often used in an informal, commonsensical sense, to mean the ability to learn and generalize beyond specific problems or contexts. Has been formalized in various ways as well, including formalizations of the notion of ”achieving complex goals in complex environments” and ”achieving complex goals in complex environments using limited resources.” Usually interpreted as a fuzzy concept, according to which absolutely general intelligence is physically unachievable, and humans have a significant level of general intelligence, but far from the maximally physically achievable degree.
- **Generalized Hypergraph**: A hypergraph with some additional features, such as links that point to links, and nodes that are seen as ”containing” whole sub-hypergraphs. This is the most natural and direct way to mathematically/visually model the Atomspace.
- **Generator**: In the PLN design, a rule is denoted a generator if it can produce its consequent without needing premises (e.g. LookupRule, which just looks it up in the AtomSpace). See composer.
- **Global, Distributed Memory**: Memory that stores items as implicit knowledge, with each memory item spread across multiple components, stored as a pattern of organization or activity among them.
- **Glocal Memory**: The storage of items in memory in a way that involves both localized and global, distributed aspects.
- **Goal**: An Atom representing a function that a system (like OpenCog) is supposed to spend a certain non-trivial percentage of its attention optimizing. The goal, informally speaking, is to maximize the Atom’s truth value.
- **Goal, Implicit**: A goal that an intelligent system, in practice, strives to achieve; but that is not explicitly represented as a goal in the system’s knowledge base.
- **Goal, Explicit**: A goal that an intelligent system explicitly represents in its knowledge base, and expends some resources trying to achieve. Goal Nodes (which may be Nodes or, e.g. ImplicationLinks) are used for this purpose in OpenCog.
- **Goal-Driven Learning**: Learning that is driven by the cognitive schematic i.e. by the quest of figuring out which procedures can be expected to achieve a certain goal in a certain sort of context.
- **Grounded SchemaNode**: See SchemaNode, Grounded.
- **Hebbian Learning**: An aspect of Attention Allocation, centered on creating and updating HebbianLinks, which represent the simultaneous importance of the Atoms joined by the HebbianLink.

- **Hebbian Links**: Links recording information about the associative relationship (co-occurrence) between Atoms. These include symmetric and asymmetric HebbianLinks.
- **Heterarchical Network**: A network of linked elements in which the semantic relationships associated with the links are generally symmetrical (e.g. they may be similarity links, or symmetrical associative links). This is one important sort of subnetwork of an intelligent system; see Dual Network.
- **Hierarchical Network**: A network of linked elements in which the semantic relationships associated with the links are generally asymmetrical, and the parent nodes of a node have a more general scope and some measure of control over their children (though there may be important feedback dynamics too). This is one important sort of subnetwork of an intelligent system; see Dual Network.
- **Higher-Order Inference (HOI)**: PLN inference involving variables or higher-order functions. In contrast to First-Order Inference (FOI).
- **Hillclimbing**: A general term for greedy, local optimization techniques, including some relatively sophisticated ones that involve "mildly nonlocal" jumps.
- **Human-Level Intelligence**: General intelligence that's "as smart as" human general intelligence, even if in some respects quite unlike human intelligence. An informal concept, which generally doesn't come up much in CogPrime work, but is used frequently by some other AI theorists.
- **Human-Like Intelligence**: General intelligence with properties and capabilities broadly resembling those of humans, but not necessarily precisely imitating human beings.
- **Hypergraph**: A conventional hypergraph is a collection of nodes and links, where each link may span any number of nodes. OpenCog makes use of generalized hypergraphs (the Atomspace is one of these).
- **Imitation Learning**: Learning via copying what some other agent is observed to do.
- **Implication**: Often refers to an ImplicationLink between two PredicateNodes, indicating an (extensional, intensional or mixed) logical implication.
- **Implicit Knowledge Representation**: Representation of knowledge via having easily humanly identifiable pieces of knowledge correspond to the pattern of organization and/or dynamics of elements, rather than via having individual elements correspond to easily humanly identifiable pieces of knowledge.
- **Importance**: A generic term for the Attention Values associated with Atoms. Most commonly these are STI (short term importance) and LTI (long term importance) values. Other importance values corresponding to various different time scales are also possible. In general an importance value reflects an estimate of the likelihood an Atom will be useful to the system over some particular future time-horizon. STI is generally relevant to processor time allocation, whereas LTI is generally relevant to memory allocation.
- **Importance Decay**: The process of Atom importance values (e.g. STI and LTI) decreasing over time, if the Atoms are not utilized. Importance decay rates may in general be context-dependent.
- **Importance Spreading**: A synonym for Importance Updating, intended to highlight the similarity with "activation spreading" in neural and semantic networks.
- **Importance Updating**: The CIM-Dynamic that periodically (frequently) updates the STI and LTI values of Atoms based on their recent activity and their relationships.
- **Imprecise Truth Value**: Peter Walley's imprecise truth values are intervals [L,U], interpreted as lower and upper bounds of the means of probability distributions in an envelope

of distributions. In general, the term may be used to refer to any truth value involving intervals or related constructs, such as indefinite probabilities.

- **Indefinite Probability**: An extension of a standard imprecise probability, comprising a credible interval for the means of probability distributions governed by a given second-order distribution.
- **Indefinite Truth Value**: An OpenCog TruthValue object wrapping up an indefinite probability
- **Induction**: In PLN, a specific inference rule (A → B, A → C, therefore B → C). In general, the process of heuristically inferring that what has been seen in multiple examples, will be seen again in new examples. Induction in the broad sense, may be carried out in OpenCog by methods other than PLN induction. When emphasis needs to be laid on the particular PLN inference rule, the phrase "PLN Induction" is used.
- **Inference**: Generally speaking, the process of deriving conclusions from assumptions. In an OpenCog context, this often refers to the PLN inference system. Inference in the broad sense is distinguished from general learning via some specific characteristics, such as the intrinsically incremental nature of inference: it proceeds step by step.
- **Inference Control**: A cognitive process that determines what logical inference rule (e.g. what PLN rule) is applied to what data, at each point in the dynamic operation of an inference process.
- **Integrative AGI**: An AGI architecture, like CogPrime, that relies on a number of different powerful, reasonably general algorithms all cooperating together. This is different from an AGI architecture that is centered on a single algorithm, and also different than an AGI architecture that expects intelligent behavior to emerge from the collective interoperation of a number of simple elements (without any sophisticated algorithms coordinating their overall behavior).
- **Integrative Cognitive Architecture**: A cognitive architecture intended to support integrative AGI.
- **Intelligence**: An informal, natural language concept. "General intelligence" is one slightly more precise specification of a related concept; "Universal intelligence" is a fully precise specification of a related concept. Other specifications of related concepts made in the particular context of CogPrime research are the pragmatic general intelligence and the efficient pragmatic general intelligence.
- **Intension**: In PLN, the intention of a node consists of Atoms representing properties of the entity the node represents.
- **Intentional memory**: A system's knowledge of its goals and their subgoals, and associations between these goals and procedures and contexts (e.g. cognitive schematics).
- **Internal Simulation World**: A simulation engine used to simulate an external environment (which may be physical or virtual), used by an AGI system as its "mind's eye" in order to experiment with various action' q sequences and envision their consequences, or observe the consequences of various hypothetical situations. Particularly important for dealing with episodic knowledge.
- **Interval Algebra**: Allen Interval Algebra, a mathematical theory of the relationships between time intervals. CogPrime utilizes a fuzzified version of classic Interval Algebra.
- **IRC Learning (Imitation, Reinforcement, Correction)**: Learning via interaction with a teacher, involving a combination of imitating the teacher, getting explicit reinforcement signals from the teacher, and having one's incorrect or suboptimal behaviors guided toward betterness by the teacher in real-time. This is a large part of how young humans learn.

- **Knowledge Base**: A shorthand for the totality of knowledge possessed by an intelligent system during a certain interval of time (whether or not this knowledge is explicitly represented). Put differently: this is an intelligence's total memory contents (inclusive of all types of memory) during an interval of time.
- **Language Comprehension**: The process of mapping natural language speech or text into a more "cognitive", largely language-independent representation. In OpenCog this has been done by various pipelines consisting of dedicated natural language processing tools, e.g. a pipeline: text → Link Parser → RelEx → RelEx2Frame → Frame2Atom Atomspace; and alternatively a pipeline Link Parser → Link2Atom → Atomspace. It would also be possible to do language comprehension purely via PLN and other generic OpenCog processes, without using specialized language processing tools.
- **Language Generation**: The process of mapping (largely language-independent) cognitive content into speech or text. In OpenCog this has been done by various pipelines consisting of dedicated natural language processing tools, e.g. a pipeline: Atomspace → NLGen → text; or more recently Atomspace → Atom2Link → surface realization → text. It would also be possible to do language generation purely via PLN and other generic OpenCog processes, without using specialized language processing tools.
- **Language Processing**: Processing of human language is decomposed, in CogPrime, into Language Comprehension, Language Generation, and Dialogue Control.
- **Learning**: In general, the process of a system adapting based on experience, in a way that increases its intelligence (its ability to achieve its goals). The theory underlying CogPrime doesn't distinguish learning from reasoning, associating, or other aspects of intelligence.
- **Learning Server**: In some OpenCog configurations, this refers to a software server that performs "offline" learning tasks (e.g. using MOSES or hillclimbing), and is in communication with an Operational Agent Controller software server that performs real-time agent control and dispatches learning tasks to and receives results from the Learning Server.
- **Linguistic Links**: A catch-all term for Atoms explicitly representing linguistic content, e.g. WordNode, SentenceNode, CharacterNode.
- **Link**: A type of Atom, representing a relationship among one or more Atoms. Links and Nodes are the two basic kinds of Atoms.
- **Link Parser**: A natural language syntax parser, created by Sleator and Temperley at Carnegie-Mellon University, and currently used as part of OpenCogPrime's natural language comprehension and natural language generation system.
- **Link2Atom**: A system for translating link parser links into Atoms. It attempts to resolve precisely as much ambiguity as needed in order to translate a given assemblage of link parser links into a unique Atom structure.
- **Lobe**: A term sometimes used to refer to a portion of a distributed Atomspace that lives in a single computational process. Often different lobes will live on different machines.
- **Localized Memory**: Memory that stores each item using a small number of closely-connected elements.
- **Logic**: In an OpenCog context, this usually refers to a set of formal rules for translating certain combinations of Atoms into "conclusion" Atoms. The paradigm case at present is the PLN probabilistic logic system, but OpenCog can also be used together with other logics.
- **Logical Links**: Any Atoms whose truth values are primarily determined or adjusted via logical rules, e.g. PLN's InheritanceLink, SimilarityLink, ImplicationLink, etc. The term isn't usually applied to other links like HebbianLinks whose semantics isn't primarily logic-

based, even though these other links can be processed via (e.g. PLN) logical inference via interpreting them logically.

- **Lojban**: A constructed human language, with a completely formalized syntax and a highly formalized semantics, and a small but active community of speakers. In principle this seems an extremely good method for communication between humans and early-stage AGI systems.
- **Lojban++**: A variant of Lojban that incorporates English words, enabling more flexible expression without the need for frequent invention of new Lojban words.
- **Long Term Importance (LTI)**: A value associated with each Atom, indicating roughly the expected utility to the system of keeping that Atom in RAM rather than saving it to disk or deleting it. It's possible to have multiple LTI values pertaining to different time scales, but so far practical implementation and most theory has centered on the option of a single LTI value.
- **LTI**: Long Term Importance
- **Map**: A collection of Atoms that are interconnected in such a way that they tend to be commonly active (i.e. to have high STI, e.g. enough to be in the AttentionalFocus, at the same time).
- **Map Encapsulation**: The process of automatically identifying maps in the Atomspace, and creating Atoms that "encapsulate" them; the Atom encapsulation a map would link to all the Atoms in the map. This is a way of making global memory into local memory, thus making the system's memory glocal and explicitly manifesting the "cognitive equation." This may be carried out via a dedicated MapEncapsulation MindAgent.
- **Map Formation**: The process via which maps form in the Atomspace. This need not be explicit; maps may form implicitly via the action of Hebbian Learning. It will commonly occur that Atoms frequently co-occurring in the AttentionalFocus, will come to be joined together in a map.
- **Memory Types**: In CogPrime
  this generally refers to the different types of memory that are embodied in different data structures or processes in the CogPrime
  architecture, e.g. declarative (semantic), procedural, attentional, intentional, episodic, sensorimotor.
- **Mind-World Correspondence Principle**: The principle that, for a mind to display efficient pragmatic general intelligence relative to a world, it should display many of the same key structural properties as that world. This can be formalized by modeling the world and mind as probabilistic state transition graphs, and saying that the categories implicit in the state transition graphs of the mind and world should be inter-mappable via a high-probability morphism.
- **Mind OS**: A synonym for the OpenCog Core.
- **MindAgent**: An OpenCog software object, residing in the CogServer, that carries out some processes in interaction with the Atomspace. A given conceptual cognitive process (e.g. PLN inference, Attention allocation, etc.) may be carried out by a number of different MindAgents designed to work together.
- **Mindspace**: A model of the set of states of an intelligent system as a geometrical space, imposed by assuming some metric on the set of mind-states. This may be used as a tool for formulating general principles about the dynamics of generally intelligent systems.
- **Modulators**: Parameters in the Psi model of motivated, emotional cognition, that modulate the way a system perceives, reasons about and interacts with the world.

- **MOSES (Meta-Optimizing Semantic Evolutionary Search)**: An algorithm for procedure learning, which in the current implementation learns programs in the Combo language. MOSES is an evolutionary learning system, which differs from typical genetic programming systems in multiple aspects including: a subtler framework for managing multiple ”demes” or ”islands” of candidate programs; a library called Reduct for placing programs in Elegant Normal Form; and the use of probabilistic modeling in place of, or in addition to, mutation and crossover as means of determining which new candidate programs to try.
- **Motoric**: Pertaining to the control of physical actuators, e.g. those connected to a robot. May sometimes be used to refer to the control of movements of a virtual character as well.
- **Moving Bubble of Attention**: The Attentional Focus of a CogPrime system.
- **Natural Language Comprehension**: See Language Comprehension
- **Natural Language Generation**: See Language Generation
- **Natural Language Processing (NLP)**: See Language Processing
- **NLGen**: Software for carrying out the surface realization phase of natural language generation, via translating collections of RelEx output relationships into English sentences. Was made functional for simple sentences and some complex sentences; not currently under active development, as work has shifted to the related Atom2Link approach to language generation.
- **Node**: A type of Atom. Links and Nodes are the two basic kinds of Atoms. Nodes, mathematically, can be thought of as "0-ary" links. Some types of Nodes refer to external or mathematical entities (e.g. WordNode, NumberNode); others are purely abstract, e.g. a ConceptNode is characterized purely by the Links relating it to other atoms. GroundedPredicateNodes and GroundedSchemaNodes connect to explicitly represented procedures (sometimes in the Combo language); ungrounded PredicateNodes and SchemaNodes are abstract and, like ConceptNodes, purely characterized by their relationships.
- **Node Probability**: Many PLN inference rules rely on probabilities associated with Nodes. Node probabilities are often easiest to interpret in a specific context, e.g. the probability P(cat) makes obvious sense in the context of a typical American house, or in the context of the center of the sun. Without any contextual specification, P(A) is taken to mean the probability that a randomly chosen occasion of the system's experience includes some instance of A.
- **Novamente Cognition Engine (NCE)**: A proprietary proto-AGI software system, the predecessor to OpenCog. Many parts of the NCE were open-sourced to form portions of OpenCog, but some NCE code was not included in OpenCog; and now OpenCog includes multiple aspects and plenty of code that was not in NCE.
- **OpenCog**: A software framework intended for development of AGI systems, and also for narrow-AI application using tools that have AGI applications. Co-designed with the CogPrime cognitive architecture, but not exclusively bound to it.
- **OpenCog Prime (OCP)**: The implementation of the CogPrime cognitive architecture within the OpenCog software framework.
- **OpenPsi**: CogPrime's architecture for motivation-driven action selection, which is based on adapting Dorner's Psi model for use in the OpenCog framework.
- **Operational Agent Controller (OAC)**: In some OpenCog configurations, this is a software server containing a CogServer devoted to real-time control of an agent (e.g. a virtual world agent, or a robot). Background, offline learning tasks may then be dispatched to other software processes, e.g. to a Learning Server.

- **Pattern**: In a CogPrime context, the term ”pattern” is generally used to refer to a process that produces some entity, and is judged simpler than that entity.
- **Pattern Mining**: Pattern mining is the process of extracting an (often large) number of patterns from some body of information, subject to some criterion regarding which patterns are of interest. Often (but not exclusively) it refers to algorithms that are rapid or ”greedy”, finding a large number of simple patterns relatively inexpensively.
- **Pattern Recognition**: The process of identifying and representing a pattern in some substrate (e.g. some collection of Atoms, or some raw perceptual data, etc.).
- **Patternism**: The philosophical principle holding that, from the perspective of engineering intelligent systems, it is sufficient and useful to think about mental processes in terms of (static and dynamical) patterns.
- **Perception**: The process of understanding data from sensors. When natural language is ingested in textual format, this is generally not considered perceptual. Perception may be taken to encompass both pre-processing that prepares sensory data for ingestion into the Atomspace, processing via specialized perception processing systems like DeSTIN that are connected to the Atomspace, and more cognitive-level process within the Atomspace that is oriented toward understanding what has been sensed.
- **Piagetan Stages**: A series of stages of cognitive development hypothesized by developmental psychologist Jean Piaget, which are easy to interpret in the context of developing CogPrime systems. The basic stages are: Infantile, Pre-operational, Concrete Operational and Formal. Post-formal stages have been discussed by theorists since Piaget and seem relevant to AGI, especially advanced AGI systems capable of strong self-modification.
- **PLN**: short for Probabilistic Logic Networks
- **PLN, First-Order**: See First-Order Inference
- **PLN, Higher-Order**: See Higher-Order Inference
- **PLN Rules**: A PLN Rule takes as input one or more Atoms (the ”premises”, usually Links), and output an Atom that is a ”logical conclusion” of those Atoms. The truth value of the consequence is determined by a PLN Formula associated with the Rule.
- **PLN Formulas**: A PLN Formula, corresponding to a PLN Rule, takes the TruthValues corresponding to the premises and produces the TruthValue corresponding to the conclusion. A single Rule may correspond to multiple Formulas, where each Formula deals with a different sort of TruthValue.
- **Pragmatic General Intelligence**: A formalization of the concept of general intelligence, based on the concept that general intelligence is the capability to achieve goals in environments, calculated as a weighted average over some fuzzy set of goals and environments.
- **Predicate Evaluation**: The process of determining the Truth Value of a predicate, embodied in a PredicateNode. This may be recursive, as the predicate referenced internally by a Grounded PredicateNode (and represented via a Combo program tree) may itself internally reference other PredicateNodes.
- **Probabilistic Logic Networks (PLN)**: A mathematical and conceptual framework for reasoning under uncertainty, integrating aspects of predicate and term logic with extensions of imprecise probability theory. OpenCogPrime’s central tool for symbolic reasoning.
- **Procedural Knowledge**: Knowledge regarding which series of actions (or action-combinations) are useful for an agent to undertake in which circumstances. In CogPrime these may be learned in a number of ways, e.g. via PLN or via Hebbian learning of Schema Maps, or via explicit learning of Combo programs via MOSES or hillclimbing. Procedures are represented as SchemaNodes or Schema Maps.

- **Procedure Evaluation/Execution**: A general term encompassing both Schema Execution and Predicate Evaluation, both of which are similar computational processes involving manipulation of Combo trees associated with ProcedureNodes.
- **Procedure Learning**: Learning of procedural knowledge, based on any method, e.g. evolutionary learning (e.g. MOSES), inference (e.g. PLN), reinforcement learning (e.g. Hebbian learning).
- **Procedure Node**: A SchemaNode or PredicateNode
- **Psi**: A model of motivated action and emotion, originated by Dietrich Dorner and further developed by Joscha Bach, who incorporated it in his proto-AGI system MicroPsi. OpenCogPrime's motivated-action component, OpenPsi, is roughly based on the Psi model.
- **Psynese**: A system enabling different OpenCog instances to communicate without using natural language, via directly exchanging Atom subgraphs, using a special system to map references in the speaker's mind into matching references in the listener's mind.
- **Psynet Model**: An early version of the theory of mind underlying CogPrime, referred to in some early writings on the Webmind AI Engine and Novamente Cognition Engine. The concepts underlying the psynet model are still part of the theory underlying CogPrime, but the name has been deprecated as it never really caught on.
- **Reasoning**: See inference
- **Reduct**: A code library, used within MOSES, applying a collection of hand-coded rewrite rules that transform Combo programs into Elegant Normal Form.
- **Region Connection Calculus**: A mathematical formalism describing a system of basic operations among spatial regions. Used in CogPrime as part of spatial inference to provide relations and rules to be referenced via PLN and potentially other subsystems.
- **Reinforcement Learning**: Learning procedures via experience, in a manner explicitly guided to cause the learning of procedures that will maximize the system's expected future reward. CogPrime does this implicitly whenever it tries to learn procedures that will maximize some Goal whose Truth Value is estimated via an expected reward calculation (where "reward" may mean simply the Truth Value of some Atom defined as "reward"). Goal-driven learning is more general than reinforcement learning as thus defined; and the learning that CogPrime does, which is only partially goal-driven, is yet more general.
- **RelEx**: A software system used in OpenCog as part of natural language comprehension, to map the output of the link parser into more abstract semantic relationships. These more abstract relationships may then be entered directly into the Atomspace, or they may be further abstracted before being entered into the Atomspace, e.g. by RelEx2Frame rules.
- **RelEx2Frame**: A system of rules for translating RelEx output into Atoms, based on the FrameNet ontology. The output of the RelEx2Frame rules make use of the FrameNet library of semantic relationships. The current (2012) RelEx2Frame rule-based is problematic and the RelEx2Frame system is deprecated as a result, in favor of Link2Atom. However, the ideas embodied in these rules may be useful; if cleaned up the rules might profitably be ported into the Atomspace as ImplicationLinks.
- **Representation Building**: A stage within MOSES, wherein a candidate Combo program tree (within a deme) is modified by replacing one or more tree nodes with alternative tree nodes, thus obtaining a new, different candidate program within that deme. This process currently relies on hand-coded knowledge regarding which types of tree nodes a given tree node should be experimentally replaced with (e.g. an AND node might sensibly be replaced with an OR node, but not so sensibly replaced with a node representing a "kick" action).

- **Request for Services (RFS)**: In CogPrime's Goal-driven action system, a RFS is a package sent from a Goal Atom to another Atom, offering it a certain amount of STI currency if it is able to deliver the goal what it wants (an increase in its Truth Value). RFS's may be passed on, e.g. from goals to subgoals to sub-subgoals, but eventually an RFS reaches a Grounded SchemaNode, and when the corresponding Schema is executed, the payment implicit in the RFS is made.
- **Robot Preschool**: An AGI Preschool in our physical world, intended for robotically embodied AGIs.
- **Robotic Embodiment**: Using an AGI to control a robot. The AGI may be running on hardware physically contained in the robot, or may run elsewhere and control the robot via networking methods such as wifi.
- **Scheduler**: Part of the CogServer that controls which processes (e.g. which MindAgents) get processor time, at which point in time.
- **Schema**: A "script" describing a process to be carried out. This may be explicit, as in the case of a GroundedSchemaNode, or implicit, as the case in Schema maps or ungrounded SchemaNodes.
- **Schema Encapsulation**: The process of automatically recognizing a Schema Map in an Atomspace, and creating a Combo (or other) program embodying the process carried out by this Schema Map, and then storing this program in the Procedure Repository and associating it with a particular SchemaNode. This translates distributed, global procedural memory into localized procedural memory. It's a special case of Map Encapsulation.
- **Schema Execution**: The process of "running" a Grounded Schema, similar to running a computer program. Or, phrased alternately: The process of executing the Schema referenced by a Grounded SchemaNode. This may be recursive, as the predicate referenced internally by a Grounded SchemaNode (and represented via a Combo program tree) may itself internally reference other Grounded SchemaNodes.
- **Schema, Grounded**: A Schema that is associated with a specific executable program (either a Combo program or, say, C++ code)
- **Schema Map**: A collection of Atoms, including SchemaNodes, that tend to be enacted in a certain order (or set of orders), thus habitually enacting the same process. This is a distributed, globalized way of storing and enacting procedures.
- **Schema, Ungrounded**: A Schema that represents an abstract procedure, not associated with any particular executable program.
- **Schematic Implication**: A general, conceptual name for implications of the form ((Context AND Procedure) IMPLIES Goal)
- **SegSim**: A name for the main algorithm underlying the NLGen language generation software. The algorithm is based on segmenting a collection of Atoms into small parts, and matching each part against memory to find, for each part, cases where similar Atom-collections already have known linguistic expression.
- **Self-Modification**: A term generally used for AI systems that can purposefully modify their core algorithms and representations. Formally and crisply distinguishing this sort of "strong self-modification" from "mere" learning is a tricky matter.
- **Sensorimotor**: Pertaining to sensory data, motoric actions, and their combination and intersection.
- **Sensory**: Pertaining to data received by the AGI system from the outside world. In a CogPrime system that perceives language directly as text, the textual input will generally

not be considered as ”sensory” (on the other hand, speech audio data would be considered as ”sensory”).

- **Short Term Importance**: A value associated with each Atom, indicating roughly the expected utility to the system of keeping that Atom in RAM rather than saving it to disk or deleting it. It’s possible to have multple LTI values pertaining to different time scales, but so far practical implementation and most theory has centered on the option of a single LTI value.
- **Similarity**: a link type indicating the probabilistic similarity between two different Atoms. Generically this is a combination of Intensional Similarity (similarity of properties) and Extensional Similarity (similarity of members).
- **Simple Truth Value**: a TruthValue involving a pair (s,d) indicating strength (e.g. probability or fuzzy set membership) and confidence d. d may be replaced by other options such as a count n or a weight of evidence w.
- **Simulation World**: See Internal Simulation World
- **SMEPH (Self-Modifying Evolving Probabilistic Hypergraphs)**: a style of modeling systems, in which each system is associated with a derived hypergraph
- **SMEPH Edge**: A link in a SMEPH derived hypergraph, indicating an empirically observed relationship (e.g. inheritance or similarity) between two
- **SMEPH Vertex**: A node in a SMEPH derived hypergraph representing a system, indicating a collection of system states empirically observed to arise in conjunction with the same external stimuli
- **Spatial Inference**: PLN reasoning including Atoms that explicitly reference spatial relationships
- **Spatiotemporal Inference**: PLN reasoning including Atoms that explicitly reference spatial and temporal relationships
- **STI**: Shorthand for Short Term Importance
- **Strength**: The main component of a TruthValue object, lying in the interval [0,1], referring either to a probability (in cases like InheritanceLink, SimilarityLink, EquivalenceLink, ImplicationLink, etc.) or a fuzzy value (as in MemberLink, EvaluationLink).
- **Strong Self-Modification**: This is generally used as synonymous with Self-Modification, in a CogPrime context.
- **Subsymbolic**: Involving processing of data using elements that have no correspondence to natural language terms, nor abstract concepts; and that are not naturally interpreted as symbolically ”standing for” other things. Often used to refer to processes such as perception processing or motor control, which are concerned with entities like pixels or commands like ”rotate servomotor 15 by 10 degrees theta and 55 degrees phi.” The distinction between ”symbolic” and ”subsymbolic” is conventional in the history of AI, but seems difficult to formalize rigorously. Logic-based AI systems are typically considered ”symbolic”, yet
- **Supercompilation**: A technique for program optimization, which globally rewrites a program into a usually very different looking program that does the same thing. A prototype supercompiler was applied to Combo programs with successful results.
- **Surface Realization**: The process of taking a collection of Atoms and transforming them into a series of words in a (usually natural) language. A stage in the overall process of language generation.
- **Symbol Grounding**: The mapping of a symbolic term into perceptual or motoric entities that help define the meaning of the symbolic term. For instance, the concept ”Cat” may be

grounded by images of cats, experiences of interactions with cats, imaginations of being a cat, etc.

- **Symbolic**: Pertaining to the formation or manipulation of symbols, i.e. mental entities that are explicitly constructed to represent other entities. Often contrasted with subsymbolic.
- **Syntax-Semantics Correlation**: In the context of MOSES and program learning more broadly, this refers to the property via which distance in syntactic space (distance between the syntactic structure of programs, e.g. if they're represented as program trees) and semantic space (distance between the behaviors of programs, e.g. if they're represented as sets of input/output pairs) are reasonably well correlated. This can often happen among sets of programs that are not too widely dispersed in program space. The Reduct library is used to place Combo programs in Elegant Normal Form, which increases the level of syntax-semantics corellation between them. The programs in a single MOSES deme are often closely enough clustered together that they have reasonably high syntax-semantics correlation.
- **System Activity Table**: An OpenCog component that records information regarding what a system did in the past.
- **Temporal Inference**: Reasoning that heavily involves Atoms representing temporal information, e.g. information about the duration of events, or their temporal relationship (before, after, during, beginning, ending). As implemented in CogPrime, makes use of an uncertain version of Allen Interval Algebra.
- **Truth Value**: A package of information associated with an Atom, indicating its degree of truth. SimpleTruthValue and IndefiniteTruthValue are two common, particular kinds. Multiple truth values associated with the same Atom from different perspectives may be grouped into CompositeTruthValue objects.
- **Universal Intelligence**: A technical term introduced by Shane Legg and Marcus Hutter, describing (roughly speaking) the average capability of a system to carry out computable goals in computable environments, where goal/environment pairs are weighted via the length of the shortest program for computing them.
- **Urge**: In OpenPsi, an Urge develops when a Demand deviates from its target range.
- **Very Long Term Importance (VLTI)**: A bit associated with Atoms, which determines whether, when an Atom is forgotten (removed from RAM), it is saved to disk (frozen) or simply deleted.
- **Virtual AGI Preschool**: A virtual world intended for AGI teaching/training/learning, bearing broad resemblance to the preschool environments used for young humans.
- **Virtual Embodiment**: Using an AGI to control an agent living in a virtual world or game world, typically (but not necessarily) a 3D world with broad similarity to the everyday human world.
- **Webmind AI Engine**: A predecessor to the Novamente Cognition Engine and OpenCog, developed 1997-2001 – with many similar concepts (and also some different ones) but quite different algorithms and software architecture

References

AABL02. Nancy Alvarado, Sam S. Adams, Steve Burbeck, and Craig Latta. Beyond the turing test: Performance metrics for evaluating a computer simulation of the human mind. *Development and Learning, International Conf. on*, 0, 2002.

AGBD$^+$08. Derek Abbott, Julio Gea-Banacloche, Paul C W Davies, Stuart Hameroff, Anton Zeilinger, Jens Eisert, Howard M. Wiseman, Sergey M. Bezrukov, and Hans Frauenfelder. Plenary debate: quantum effects in biology?trivial or not? *Fluctuation and Noise Letters 8(1), pp. C5ĐC26*, 2008.

AL03. J. R. Anderson and C. Lebiere. The newell test for a theory of cognition. *Behavioral and Brain Science*, 26, 2003.

AL09. Itamar Arel and Scott Livingston. Beyond the turing test. *IEEE Computer*, 42(3):90–91, March 2009.

AM01. J. S. Albus and A. M. Meystel. *Engineering of Mind: An Introduction to the Science of Intelligent Systems*. Wiley and Sons, 2001.

Ami89. Daniel J. Amit. *Modeling brain function – the world of attractor neural networks*. Cambridge University Press, New York, USA, 1989.

ARC09. I. Arel, D. Rose, and R. Coop. Destin: A scalable deep learning architecture with application to high-dimensional robust pattern recognition. *Proc. AAAI Workshop on Biologically Inspired Cognitive Architectures*, 2009.

ARK09a. I. Arel, D. Rose, and T. Karnowski. A deep learning architecture comprising homogeneous cortical circuits for scalable spatiotemporal pattern inference. *NIPS 2009 Workshop on Deep Learning for Speech Recognition and Related Applications*, 2009.

Ark09b. Ronald Arkin. *Governing Lethal Behavior in Autonomous Robots*. Chapman and Hall, 2009.

Arl75. P. K. Arlin. *Cognitive development in adulthood: A fifth stage?*, volume 11. Developmental Psychology, 1975.

Arm04. J. Andrew Armour. Cardiac neuronal hierarchy in health and disease. *Am J Physiol Regul Integr Comp Physiol 287:*, 2004.

Baa97. Bernard Baars. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press, 1997.

Bac09. Joscha Bach. *Principles of Synthetic Intelligence*. Oxford University Press, 2009.

Bar02. Albert-Laszlo Barabasi. *Linked: The New Science of Networks*. Perseus, 2002.

Bat79. Gregory Bateson. *Mind and Nature: A Necessary Unity*. New York: Ballantine, 1979.

BC94. S. Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, 1994.

BDL93. Louise Barrett, Robin Dunbar, and John Lycett. *Human Evolutionary Psychology*. Princeton University Press, 1993.

BDS03. S Ben-David and R Schuller. Exploiting task relatedness for learning multiple tasks. *Proceedings of the 16th Annual Conference on Learning Theory*, 2003.

BF71. J. D. Bransford and J. Franks. The abstraction of linguistic ideas. *Cognitive Psychology*, 2:331–350, 1971.

BF09. Bernard Baars and Stan Franklin. Consciousness is computational: The lida model of global workspace theory. *International Journal of Machine Consciousness.*, 2009.

bGBK02.

1. Goertzel, Andrei Klimov Ben, and Arkady Klimov. Supercompiling java programs, 2002.

BH05. Sebastian Bader and Pascal Hitzler. Dimensions of neural-symbolic integration - a structured survey. In S. Artemov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb, and J. Woods., editors, *We Will Show Them: Essays in Honour of Dov Gabbay*, volume 1, pages 167–194. College Publications, 2005.

Bi01. M-m Bi, G-q andPoo. Synaptic modifications by correlated activity: Hebb's postulate revisited. *Ann Rev Neurosci ; 24:139-166*, 2001.

Bic88. M. Bickhard. Piaget on variation and selection models: Structuralism, logical necessity, and interactivism. *Human Development*, 31:274–312, 1988.

Bil05. Philip Bille. A survey on tree edit distance and related problems. *Theoretical Computer Science*, 337:2005, 2005.

BO09. A. Baranes and Pierre-Yves Oudeyer. R-iac: Robust intrinsically motivated active learning. *Proc. of the IEEE International Conf. on Learning and Development, Shanghai, China.*, 33, 2009.

Bol98. B. Bollobas. *Modern Graph Theory*. Springer, 1998.

Bos02. Nick Bostrom. Existential risks. *Journal of Evolution and Technology*, 9, 2002.

Bos03. Nick Bostrom. Ethical issues in advanced artificial intelligence. In Iva Smit, editor, *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, volume 2., pages 12–17. 2003.

Bro84. J. Broughton. Not beyond formal operations, but beyond piaget. In M. Commons, F. Richards, and C. Armon, editors, *Beyond Formal Operations: Late Adolescent and Adult Cognitive Development*, pages 395–411. Praeger. New York, 1984.

BS04. B. Bakker and Juergen Schmidhuber. Hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization. *Proc. of the 8-th Conf. on Intelligent Autonomous Systems*, 2004.

Buc03. Mark Buchanan. *Small World: Uncovering Nature's Hidden Networks.* Phoenix, 2003.

Bur62. C MacFarlane Burnet. *The Integrity of the Body.* Harvard University Press, 1962.

BW88. R. W. Byrne and A. Whiten. *Machiavellian Intelligence.* Clarendon Press, 1988.

BZ03. Selmer Bringsjord and M Zenzen. *Superminds: People Harness Hypercomputation, and More.* Kluwer, 2003.

BZGS06. B. Bakker, V. Zhumatiy, G. Gruener, and Juergen Schmidhuber. Quasi-online reinforcement learning for robots. *Proc. of the International Conf. on Robotics and Automation*, 2006.

Cal96. William Calvin. *The Cerebral Code.* MIT Press, 1996.

Car85. S. Carey. *Conceptual Change in Childhood.* MIT Press, 1985.

Car97. R Caruana. Multitask learning. *Machine Learning*, 1997.

Cas85. R. Case. *Intellectual development: Birth to adulthood.* Academic Press, 1985.

Cas04. N. L. Cassimatis. Grammatical processing using the mechanisms of physical inferences. In *Proceedings of the Twentieth-Sixth Annual Conference of the Cognitive Science Society.* 2004.

Cas07. Nick Cassimatis. Adaptive algorithmic hybrids for human-level artificial intelligence. 2007.

CB00. W. H. Calvin and D. Bickerton. *Lingua ex Machina.* MIT Press, 2000.

CB06. Rory Conolly and Jerry Blancato. Computational modeling of the liver. *NCCT BOSC Review*, 2006. http://www.epa.gov/ncct/bosc_review/2006/files/07_Conolly_Liver_Model.pdf.

CM07. Jie-Qi Chen and Gillian McNamee. *What is Waldorf Education?* Bridging: Assessment for Teaching and Learning in Early Childhood Classrooms, 2007.

CP05. M. L. Commons and A. Pekker. Hierarchical complexity: A formal theory. http://www.dareassociation.org/Papers/Hierarchical%20Complexity%20-%20A%20Formal%20Theory%20(Commons%20&%20Pekker).pdf, 2005.

CRK82. M. Commons, F. Richards, and D. Kuhn. Systematic and metasystematic reasoning: a case for a level of reasoning beyond Piaget's formal operations. *Child Development*, 53.:1058–1069, 1982.

CS90. A. G. Cairns-Smith. *Seven Clues to the Origin of Life: A Scientific Detective Story.* Cambridge University Press, 1990.

Cse06. Peter Csermely. *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks.* Springer, 2006.

CSG07. Subhojit Chakraborty, Anders Sandberg, and Susan A Greenfield. Differential dynamics of transient neuronal assemblies in visual compared to auditory cortex. *Experimental Brain Research, 1432-1106*, 2007.

CTS$^+$98. M. Commons, E. J. Trudeau, S. A. Stein, F. A. Richards, and S. R. Krause. Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review. 18*, 18.:237–278, 1998.

Dam00. Antonio Damasio. *The Feeling of What Happens.* Harvest Books, 2000.

Dav84. D. Davidson. *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press, 1984.

DC02. Roberts P D and Bell C C. Spike-timing dependent synaptic plasticity in biological systems. *Biological Cybernetics, 87, 392-403*, 2002.

Den87. D. Dennett. *The Intentional Stance.* Cambridge, MA: MIT Press, 1987.

Den91. Daniel Dennett. *Consciousness Explained.* Back Bay, 1991.

DG05. Hugo De Garis. *The Artilect War.* ETC, 2005.

DOP08. Wlodzislaw Duch, Richard Oentaryo, and Michel Pasquier. Cognitive architectures: Where do we go from here? *Proc. of the Second Conf. on AGI*, 2008.

Dör02. Dietrich Dörner. *Die Mechanik des Seelenwagens. Eine neuronale Theorie der Handlungsregulation.* Verlag Hans Huber, 2002.

EBJ$^+$97. J. Elman, E. Bates, M. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. *Rethinking Innateness: A Connectionist Perspective on Development.* MIT Press, 1997.

Ede93. Gerald Edelman. Neural darwinism: Selection and reentrant signaling in higher brain function. *Neuron*, 10, 1993.

Elm91. J. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–226, 1991.

EMC12. Effective-Mind-Control.com. Cellular memory in organ transplants. *Effective Mind Control,*, 2012. http://www.effective-mind-control.com/cellular-memory-in-organ-transplants.html, updated Feb 1 2012.

ES00. G. Engelbretsen and F. Sommers. *An invitation to formal reasoning. The Logic of Terms.* Aldershot: Ashgate, 2000.

FB08. Stan Franklin and Bernard Baars. Possible neural correlates of cognitive processes and modules from the lida model of cognition. *Cognitive Computing Research Group, University of Memphis*, 2008. http://ccrg.cs.memphis.edu/tutorial/correlates.html.

FC86. R. Fung and C. Chong. Metaprobability and Dempster-shafer in evidential reasoning. In L. Kanal and J. Lemmer. North-Holland, editors, *Uncertainty in Artificial Intelligence*, pages 295–302. 1986.

Fis80. K. Fischer. A theory of cognitive development: control and construction of hierarchies of skills. *Psychological Review*, 87:477–531, 1980.

Fis01. Jefferson M. Fish. *Race and Intelligence: Separating Science From Myth.* Routledge, 2001.

Fod94. J. Fodor. *The Elm and the Expert.* Cambridge, MA: Bradford Books, 1994.

FP86. Doyne Farmer and Alan Perelson. The immune system, adaptation and machine learning. *Physica D, v. 2*, 1986.

Fra06. Stan Franklin. The lida architecture: Adding new modes of learning to an intelligent, autonomous, software agent. *Int. Conf. on Integrated Design and Process Technology*, 2006.

Fre90. R. French. Subcognition and the limits of the turing test'. *Mind*, 1990.

Fre95. Walter Freeman. *Societies of Brains.* Erlbaum, 1995.

FT02. G. Fauconnier and M. Turner. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* Basic, 2002.

Gar99. H Gardner. *Intelligence reframed: Multiple intelligences for the 21st century.* Basic, 1999.

GD09. Ben Goertzel and Deborah Duong. Opencog ns: An extensible, integrative architecture for intelligent humanoid robotics. 2009.

GdG08. Ben Goertzel and Hugo de Garis. Xia-man: An extensible, integrative architecture for intelligent humanoid robotics. pages 86–90, 2008.

GE86. R. Gelman and E. Meck and s. *Merkin (1986). Young children's numerical competence. Cognitive Development*, 1:1–29, 1986.

GEA08. Ben Goertzel and Cassio Pennachin Et Al. An integrative methodology for teaching embodied non-linguistic agents, applied to virtual animals in second life. In *Proc.of the First Conf. on AGI.* IOS Press, 2008.

Ger99. Michael Gershon. *The Second Brain.* Harper, 1999.

GGC^{+}11. Ben Goertzel, Nil Geisweiller, Lucio Coelho, Predrag Janicic, and Cassio Pennachin. *Real World Reasoning.* Atlantis, 2011.

GGK02. T. Gilovich, D. Griffin, and D. Kahneman. *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press, 2002.

Gib77. J. J. Gibson. The theory of affordances. In R. Shaw & J. Bransford. Erlbaum, editor, *Perceiving, Acting and Knowing.* 1977.

Gib78. John Gibbs. Kohlberg's moral stage theory: a Piagetian revision. *Human Development*, 22:89–112, 1978.

Gib79. J. J. Gibson. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin, 1979.

GIGH08. B. Goertzel, M. Ikle, I. Goertzel, and A. Heljakka. *Probabilistic Logic Networks.* Springer, 2008.

Gil82. Carol Gilligan. *In a Different Voice.* Cambridge, MA: Harvard University Press, 1982.

GMIH08. B. Goertzel, I. Goertzel M. Iklé, and A. Heljakka. *Probabilistic Logic Networks.* Springer, 2008.

Goe93a. Ben Goertzel. *The Evolving Mind.* Plenum, 1993.

Goe93b. Ben Goertzel. *The Structure of Intelligence.* Springer, 1993.

Goe94. Ben Goertzel. *Chaotic Logic.* Plenum, 1994.

Goe97. Ben Goertzel. *From Complexity to Creativity.* Plenum Press, 1997.

Goe01. Ben Goertzel. *Creating Internet Intelligence.* Plenum Press, 2001.

Goe06a. Ben Goertzel. *The Hidden Pattern.* Brown Walker, 2006.

Goe06b. Ben Goertzel. *The Hidden Pattern.* Brown Walker, 2006.

Goe08. Ben Goertzel. A pragmatic path toward endowing virtually-embodied ais with human-level linguistic capability. IEEE World Congress on Computational Intelligence (WCCI), 2008.

Goe09a. Ben Goertzel. Cognitive synergy: A universal principle of feasible general intelligence? In *ICCI 2009, Hong Kong*, 2009.

Goe09b. Ben Goertzel. The embodied communication prior. In *Proceedings of ICCI-09, Hong Kong*, 2009.

Goe09c. Ben Goertzel. Opencog prime: A cognitive synergy based architecture for embodied artificial general intelligence. In *ICCI 2009, Hong Kong*, 2009.

Goe10a. Ben Goertzel. Coherent aggregated volition. *Multiverse According to Ben*, 2010. `http://multiverseaccordingtoben.blogspot.com/2010/03/coherent-aggregated-volition-toward.htm`.

Goe10b. Ben Goertzel. Opencogprime wikibook. 2010. `http://wiki.opencog.org/w/OpenCogPrime:WikiBook`.

Goe10c. Ben Goertzel. Toward a formal definition of real-world general intelligence. 2010.

Goe10d. Ben et al Goertzel. A general intelligence oriented architecture for embodied natural language processing. In *Proc. of the Third Conf. on Artificial General Intelligence (AGI-10)*. Atlantis Press, 2010.

Goo86. I. Good. *The Estimation of Probabilities*. Cambridge, MA: MIT Press, 1986.

Gor86. R. Gordon. Folk psychology as simulation. *Mind and Language. 1*, 1.:158–171, 1986.

GPC^{+}11. Ben Goertzel, Joel Pitt, Zhenhua Cai, Jared Wigmore, Deheng Huang, Nil Geisweiller, Ruiting Lian, and Gino Yu. Integrative general intelligence for controlling game ai in a minecraft-like environment. In *Proc. of BICA 2011*, 2011.

GPI^{+}10. Ben Goertzel, Joel Pitt, Matthew Ikle, Cassio Pennachin, and Rui Liu. Glocal memory: a design principle for artificial brains and minds. *Neurocomputing*, April 2010.

GPPG06. Ben Goertzel, Hugo Pinto, Cassio Pennachin, and Izabela Freire Goertzel. Using dependency parsing and probabilistic inference to extract relationships between genes, proteins and malignancies implicit among multiple biomedical research abstracts. In *Proc. of Bio-NLP 2006*, 2006.

GPSL03. Ben Goertzel, Cassio Pennachin, Andre' Senna, and Moshe Looks. An integrative architecture for artificial general intelligence. In *Proceedings of IJCAI 2003, Acapulco*, 2003.

Gre01. Susan Greenfield. *The Private Life of the Brain*. Wiley, 2001.

GRM^{+}11. Erik M. Gauger, Elisabeth Rieper, John J. L. Morton, Simon C. Benjamin, and Vlatko Vedral. Sustained quantum coherence and entanglement in the avian compass. *Physics Review Letters, vol. 106, no. 4*, 2011.

HAG07. Markert H, Knoblauch A, and Palm G. Modelling of syntactical processing in the cortex. *Biosystems May-Jun; 89(1-3): 300-15*, 2007.

Ham87. Stuart Hameroff. *Ultimate Computing*. North Holland, 1987.

Ham10. Stuart Hameroff. The Òconscious pilotÓÑdendritic synchrony moves through the brain to mediate consciousness. *Journal of Biological Physics*, 2010.

Hay85. Patrick Hayes. The second naive physics manifesto. In R. Shaw & J. Bransford, editor, *Formal Theories of the Commonsense World*. 1985.

HB06. Jeff Hawkins and Sandra Blakeslee. *On Intelligence*. Brown Walker, 2006.

Heb49. Donald Hebb. *The organization of behavior*. Wiley, 1949.

Hey07. F. Heylighen. *The Global Superorganism: an evolutionary-cybernetic model of the emerging network society*. Social Evolution and History 6-1, 2007.

HF95. P. Hayes and K. Ford. Turing test considered harmful. *IJCAI-14*, 1995.

HG08. David Hart and Ben Goertzel. Opencog: A software framework for integrative artificial general intelligence. In *AGI*, volume 171 of *Frontiers in Artificial Intelligence and Applications*, pages 468–472. IOS Press, 2008.

HHPO12. Adam Hampshire, Roger Highfield, Beth Parkin, and Adrian Owen. Fractionating human intelligence. *Neuron vol. 76 issue 6*, 2012.

Hib02. Bill Hibbard. *Superintelligent Machines*. Springer, 2002.

Hof79. Douglas Hofstadter. *Godel, Escher, Bach: An Eternal Golden Braid*. Basic, 1979.

Hof95. Douglas Hofstadter. *Fluid Concepts and Creative Analogies*. Basic Books, 1995.

Hof96. Douglas Hofstadter. *Metamagical Themas*. Basic Books, 1996.

Hop82. J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. of the National Academy of Sciences*, 79:2554–2558, 1982.

HOT06. G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

Hut95. E. Hutchins. *Cognition in the Wild.* MIT Press, 1995.

Hut96. Edwin Hutchins. *Cognition in the Wild.* MIT Press, 1996.

Hut05. Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability.* Springer, 2005.

HZT^{+}02. J. Han, S. Zeng, K. Tham, M. Badgero, and J. Weng. Dav: A humanoid robot platform for autonomous mental development,. *Proc. 2nd International Conf. on Development and Learning*, 2002.

IP58. B. Inhelder and J. Piaget. *The Growth of Logical Thinking from Childhood to Adolescence.* Basic Books, 1958.

JL08. D. J. Jilk and C. Lebiere. and o'reilly. *R. C. and Anderson, J. R. (2008). SAL: An explicitly pluralistic cognitive architecture. Journal of Experimental and Theoretical Artificial Intelligence*, 20:197–218, 2008.

JM09. Daniel Jurafsky and James Martin. *Speech and Language Processing.* Pearson Prentice Hall, 2009.

Joy00. Bill Joy. *Why the future doesn't need us, Wired.* April 2000.

Kam91. George Kampis. *Self-Modifying Systems in Biology and Cognitive Science.* Plenum Press, 1991.

Kan64. Immanuel Kant. *Groundwork of the Metaphysic of Morals.* Harper and Row, 1964.

Kap08. F. Kaplan. Neurorobotics: an experimental science of embodiment. *Frontiers in Neuroscience*, 2008.

KE06. J. L. Krichmar and G. M. Edelman. Principles underlying the construction of brain-based devices. In T. Kovacs and J. A. R. Marshall, editors, *Adaptation in Artificial and Biological Systems*, pages 37–42. 2006.

KK90. K. Kitchener and P. King. Reflective judgement: ten years of research. In M. Commons. Praeger. New York, editor, *Beyond Formal Operations: Models and Methods in the Study of Adolescent and Adult Thought*, volume 2, pages 63–78. 1990.

KLH83. Lawrence Kohlberg, Charles Levine, and Alexandra Hewer. *Moral stages : a current formulation and a response to critics.* Karger. Basel, 1983.

Koh38. Wolfgang Kohler. *The Place of Value in a World of Facts.* Liveright Press, New York, 1938.

Koh81. Lawrence Kohlberg. *Essays on Moral Development*, volume I. The Philosophy of Moral Development, 1981.

KS04. Adam Kahane and Peter Senge. *Solving Tough Problems: An Open Way of Talking, Listening, and Creating New Realities.* Berrett-Koehler, 2004.

Kur06. Ray Kurzweil. *The Singularity is Near.* 2006.

Kur12. Ray Kurzweil. *How to Create a Mind.* Viking, 2012.

Kyb97. H. Kyburg. Bayesian and non-bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1997.

Lan05. Pat Langley. An adaptive architecture for physical agents. *Proc. of the 2005 IEEE/WIC/ACM Int. Conf. on Intelligent Agent Technology*, 2005.

LAon. C. Lebiere and J. R. Anderson. The case for a hybrid architecture of cognition. (in preparation).

LBDE90. Y. LeCun, B. Boser, J. S. Denker, and Al. Et. Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, 2, 1990.

LD03. A. Laud and G. Dejong. The influence of reward on the speed of reinforcement learning. *Proc. of the 20th International Conf. on Machine Learning*, 2003.

Leg06a. Shane Legg. Friendly ai is bunk. *Vetta Project*, 2006. `http://commonsenseatheism.com/wp-content/uploads/2011/02/Legg-Friendly-AI-is-bunk.pdf`.

Leg06b. Shane Legg. Unprovability of friendly ai. *Vetta Project*, 2006. `http://www.vetta.org/2006/09/unprovability-of-friendly-ai/`.

LG90. Douglas Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project.* Addison-Wesley, 1990.

LH07a. Shane Legg and Marcus Hutter. A collection of definitions of intelligence. IOS, 2007.

LH07b. Shane Legg and Marcus Hutter. A definition of machine intelligence. *Minds and Machines*, 17, 2007.

LLW^{+}05. Guang Li, Zhengguo Lou, Le Wang, Xu Li, and Walter J Freeman. Application of chaotic neural model based on olfactory system on pattern recognition. *ICNC*, 1:378–381, 2005.

LMC07a. M. H. Lee, Q. Meng, and F. Chao. Developmental learning for autonomous robots. *Robotics and Autonomous Systems*, 2007.

LMC07b. M. H. Lee, Q. Meng, and F. Chao. Staged competence learning in developmental robotics. *Adaptive Behavior*, 2007.

LN00. George Lakoff and Rafael Nunez. *Where Mathematics Comes From.* Basic Books, 2000.

Log07. Robert M. Logan. *The Extended Mind.* University of Toronto Press, 2007.

Loo06. Moshe Looks. *Competent Program Evolution.* PhD Thesis, Computer Science Department, Washington University, 2006.

LRN87. John Laird, Paul Rosenbloom, and Alan Newell. Soar: An architecture for general intelligence. *Artificial Intelligence*, 33, 1987.

LS05. J Lisman and N Spruston. Postsynaptic depolarization requirements for ltp and ltd: a critique of spike timing-dependent plasticity. *Nature Neuroscience 8, 839-41*, 2005.

LWML09. John Laird, Robert Wray, Robert Marinier, and Pat Langley. Claims and challenges in evaluating human-level intelligent systems. *Proc. of AGI-09*, 2009.

Mac95. D. MacKenzie. The automation of proof: A historical and sociological exploration. *IEEE Annals of the History of Computing*, 17(3):7–29, 1995.

Mar01. H. Marchand. *Reflections on PostFormal Thought.* The Genetic Epistemologist, 2001.

McK03. Bill McKibben. *Enough: Staying Human in an Engineered Age.* Saint Martins Griffin, 2003.

Met04. Thomas Metzinger. *Being No One.* Bradford, 2004.

Min88. Marvin Minsky. *The Society of Mind.* MIT Press, 1988.

Min07. Marvin Minsky. *The Emotion Machine.* 2007.

MK07. Joseph Modayil and Benjamin Kuipers. Autonomous development of a grounded object ontology by a learning robot. *AAAI-07*, 2007.

MK08. Jonathan Mugan and Benjamin Kuipers. Towards the application of reinforcement learning to undirected developmental learning. *International Conf. on Epigenetic Robotics*, 2008.

MK09. Jonathan Mugan and Benjamin Kuipers. Autonomously learning an action hierarchy using a learned qualitative state representation. *IJCAI-09*, 2009.

Mon12. Maria Montessori. *The Montessori Method.* Frederick A. Stokes, 1912.

MSV$^+$08. G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The icub humanoid robot: an open platform for research in embodied cognition. *Performance Metrics for Intelligent Systems Workshop (PerMIS 2008)*, 2008.

MW07. Stephen Morgan and Christopher Winship. *Counterfactuals and Causal Inference.* Cambridge University Press, 2007.

Nan08. Nanowerk. Carbon nanotube rubber could provide e-skin for robots. `http://www.nanowerk.com/news/newsid=6717.php`, 2008.

Nei98. Dianne Miller Neilsen. *Teaching Young Children, Preschool-K: A Guide to Planning Your Curriculum, Teaching Through Learning Centers, and Just About Everything Else.* Corwin Press, 1998.

New90. Alan Newell. *Unified Theories of Cognition.* Harvard University press, 1990.

Nie98. Dianne Miller Nielsen. *Teaching Young Children, Preschool-K: A Guide to Planning Your Curriculum, Teaching Through Learning Centers, and Just About Everything Else.* Corwin Press, 1998.

Nil09. Nils Nilsson. The physical symbol system hypothesis: Status and prospects. *50 Years of AI, Festschrift, LNAI 4850*, 33, 2009.

NK04. A. Nestor and B. Kokinov. Towards active vision in the dual cognitive architecture. *International Journal on Information Theories and Applications*, 11, 2004.

OK06. P. Oudeyer and F. Kaplan. Discovering communication. *Connection Science*, 2006.

Omo08. Stephen Omohundro. The basic ai drives. Proceedings of the First AGI Conference. IOS Press, 2008.

Omo09. Stephen Omohundro. Creating a cooperative future. 2009. `http://selfawaresystems.com/2009/02/23/talk-on-creating-a-cooperative-future/`.

Opa52. A. I. Oparin. *The Origin of Life.* Dover, 1952.

Pal82. Gunter Palm. *Neural Assemblies. An Alternative Approach to Artificial Intelligence.* Springer, 1982.

Pei34. C. Peirce. *Collected papers: Volume V. Pragmatism and pragmaticism.* Harvard University Press. Cambridge MA., 1934.

Pel05. Martin Pelikan. *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms.* Springer, 2005.

Pen96. Roger Penrose. *Shadows of the Mind.* Oxford University Press, 1996.

Per70. William G. Perry. *Forms of Intellectual and Ethical Development in the College Years: A Scheme.* Holt, Rinehart and Winston, 1970.

Per81. William G. Perry. Cognitive and ethical growth: The making of meaning. In Arthur W. Chickering. Jossey-Bass. San Francisco, editor, *The Modern American College*, pages 76–116. 1981.

PH12. Zhiping Pang and Weiping Han. Regulation of synaptic functions in central nervous system by endocrine hormones and the maintenance of energy homeostasis. *Bioscience Reports*, 2012.

Pia53. Jean Piaget. *The Origins of Intelligence in Children.* Routledge and Kegan Paul, 1953.

Pia55. Jean Piaget. *The Construction of Reality in the Child.* Routledge and Kegan Paul, 1955.

Pir84. Robert Pirsig. *Zen and the Art of Motorcycle Maintenance.* Bantam, 1984.

PNR07. Karalny Patterson, Peter J. Nestor, and Timothy T. Rogers. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8:976–987, 2007.

PSF09. Richard Dum Peter Strick and Julie Fiez. Cerebellum and nonmotor function. *Annual Review of Neuroscience Vol. 32: 413-434*, 2009.

PW78. D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, pages 515–526, 1978.

QaGKKF05. R. Quian Quiroga, L. Reddy amd G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single-neurons in the human brain. *Nature*, 435:1102–1107, 2005.

QKKF08. R. Quian Quiroga, G Kreiman, C Koch, and I. Fried. Sparse but not "grandmother-cell" coding in the medial temporal lobe. *Trends in Cognitive Sciences*, 12:87–91, 2008.

Rav04. Ian Ravenscroft. Folk psychology as a theory, stanford encyclopedia of philosophy. `http://plato.stanford.edu/entries/folkpsych-theory/`, 2004.

RBW92. Gagne R., L. Briggs, and W. Walter. *Principles of Instructional Design.* Harcourt Brace Jovanovich, 1992.

RCK01. J. Rosbe, R. S. Chong, and D. E. Kieras. Modeling with perceptual and memory constraints: An epic-soar model of a simplified enroute air traffic control task. *SOAR Technology Inc. Report*, 2001.

RD06. Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 2006.

Rie73. K. Riegel. Dialectic operations: the final phase of cognitive development. *Human Development*, 16.:346–370, 1973.

RM95. H. L. Roediger and K. B. McDermott. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:803–814, 1995.

Ros88. Israel Rosenfield. *The Invention of Memory: A New View of the Brain.* Basic Books, 1988.

Row90. John Rowan. *Subpersonalities: The People Inside Us.* Routledge Press, 1990.

Row11. T Rowe. Fossil evidence on origin of the mammalian brain. *Science 20*, 2011.

RV01. Alan Robinson and Andrei Voronkov. *Handbook of Automated Reasoning.* MIT Press, 2001.

RZDK05. Michael Rosenstein, ZvikaMarx, Tom Dietterich, and Leslie Pack Kaelbling. Transfer learning with an ensemble of background tasks. *NIPS workshop on inductive transfer*, 2005.

SA93. L. Shastri and V. Ajjanagadde. From simple associations to systematic reasoning: A connectionist encoding of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences*, 16-3, 1993.

Sal93. Stan Salthe. *Development and Evolution.* MIT Press, 1993.

Sam10. Alexei V. Samsonovich. Toward a unified catalog of implemented cognitive architectures. In *BICA*, pages 195–244, 2010.

SB98. Richard Sutton and Andrew Barto. *Reinforcement Learning.* MIT Press, 1998.

SB06. J. Simsek and A. Barto. An intrinsic reward mechanism for efficient exploration. *Proc. of the Twenty-Third International Conf. on Machine Learning*, 2006.

SBC05. S. Singh, A. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. *Proc. of Neural Information Processing Systems 17*, 2005.

SC94. Barry Smith and Roberto Casati. *Naive Physics: An Essay in Ontology.* Philosophical Psychology, 1994.

Sch91a. Juergen Schmidhuber. Curious model-building control systems.. *Proc. International Joint Conf. on Neural Networks*, 1991.

Sch91b. Juergen Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. *Proc. of the International Conf. on Simulation of Adaptive Behavior: From Animals to Animats*, 1991.

Sch95. Juergen Schmidhuber. Reinforcement-driven information acquisition in non-deterministic environments. *Proc. ICANN'95*, 1995.

Sch02. Juergen Schmidhuber. Exploring the predictable.. Springer, 2002.

Sch06. J. Schmidhuber. Godel machines: Fully Self-referential Optimal Universal Self-improvers. In B. Goertzel and C. Pennachin, editors, *Artificial General Intelligence*, pages 119–226. 2006.

Sch07. Dale Schunk. *Theories of Learning: An Educational Perspective.* Prentice Hall, 2007.

SE07. Stuart Shapiro and Al. Et. Metacognition in sneps. *AI Magazine*, 28, 2007.

SF05. Greenfield SA and Collins T F. A neuroscientific approach to consciousness. *Prog Brain Res.*, 2005.

Sha76. G. Shafer. *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press, 1976.

Shu03. Thomas R. Shultz. *Computational Developmental Psychology.* MIT Press, 2003.

SKBB91. D Shannahoff-Khalsa, M Boyle, and M Buebel. The effects of unilateral forced nostril breathing on cognition. *Int J Neurosci.*, 1991.

Slo01. Aaron Sloman. Varieties of affect and the cogaff architecture schema. In *Proceedings of the Symposium on Emotion, Cognition, and Affective Computing*, AISB-01, 2001.

Slo08a. Aaron Sloman. A new approach to philosophy of mathematics: Design a young explorer able to discover 'toddler theorems'. 2008.

Slo08b. Aaron Sloman. *The Well-Designed Young Mathematician.* Artificial Intelligence, December 2008.

SM05. Push Singh and Marvin Minsky. An architecture for cognitive diversity. In Darryl Davis, editor, *Visions of Mind*. 2005.

Sot11. Kaj Sotala. 14 objections against ai/friendly ai/the singularity answered. *Xuenay.net*, 2011. `http://www.xuenay.net/objections.html` , downloaded 3/20/11.

SS74. Jean Sauvy and Simonne Suavy. *The Child's Discovery of Space: From hopscotch to mazes – an introduction to intuitive topology.* Penguin, 1974.

SS03a. John F. Santore and Stuart C. Shapiro. Crystal cassie: Use of a 3-d gaming environment for a cognitive agent. In *Papers of the IJCAI 2003 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, 2003.

SS03b. Rudolf Steiner and S K Sagarin. *What is Waldorf Education?* Steiner Books, 2003.

Stc00. Theodore Stcherbatsky. *Buddhist Logic.* Motilal Banarsidass Pub, 2000.

SV99. A. J. Storkey and R. Valabregue. The basins of attraction of a new hopfield learning rule. *Neural Networks*, 12:869–876, 1999.

SZ04. R. Sun and X. Zhang. Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research*, 5, 2004.

TC97. M. Tomasello and J. Call. *Primate Cognition.* Oxford University Press, 1997.

TC05. Endel Tulving and R. Craik. *The Oxford Handbook of Memory.* Oxford U. Press, 2005.

Tea06. Sebastian Thrun and et al. The robot that won the darpa grand challenge. *Journal of Robotic Systems*, 23-9, 2006.

TM95. S. Thrun and Tom Mitchell. Lifelong robot learning. *Robotics and Autonomous Systems*, 1995.

TS94. E. Thelen and L. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action.* MIT Press, 1994.

TS07. M. Taylor and P. Stone. Cross-domain transfer for reinforcement learning. *Proc. of the 24th International Conf. on Machine Learning*, 2007.

Tur50. Alan Turing. Computing machinery and intelligence. *Mind*, 59, 1950.

Tur77. Valentin F. Turchin. *The Phenomenon of Science.* Columbia University Press, 1977.

TV96. Turchin and V. Supercompilation: Techniques and results. In Dines Bjorner, M. Broy, and Aleksandr Vasilevich Zamulin, editors, *Perspectives of System Informatics.* Springer, 1996.

Vin93. Vernor Vinge. The coming technological singularity. *VISION-21 Symposium, NASA and Ohio Aerospace Institute*, 1993. `http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html`.

Vyg86. Lev Vygotsky. *Thought and Language.* MIT Press, 1986.

WA10. Wendell Wallach and Colin Atkins. *Moral Machines.* Oxford University Press, 2010.

Wan95. P. Wang. *Non-Axiomatic Reasoning System.* PhD Thesis, Indiana University. Bloomington, 1995.

Wan06. Pei Wang. *Rigid Flexibility: The Logic of Intelligence.* Springer, 2006.

Was09. Mark Waser. Ethics for self-improving machines. In *AGI-09*, 2009. `http://vimeo.com/3698890`.

Wel90. H. Wellman. *The Child's Theory of Mind.* MIT Press, 1990.

WH06. J. Weng and W. S. Hwangi. From neural networks to the brain: Autonomous mental development. *IEEE Computational Intelligence Magazine*, 2006.

Who64. Benjamin Lee Whorf. *Language, Thought and Reality.* 1964.

WHZ+00. J. Weng, W. S. Hwang, Y. Zhang, C. Yang, and R. Smith. Developmental humanoids: Humanoids that develop skills automatically,. *Proc. the first IEEE-RAS International Conf. on Humanoid Robots*, 2000.

Wik11. Wikipedia. Open source governance. 2011. http://en.wikipedia.org/wiki/Open_source_governance.

Win72. Terry Winograd. *Understanding Natural Language*. Edinburgh University Press, 1972.

Wit07. David C. Witherington. *The Dynamic Systems Approach as Metatheory for Developmental Psychology, Human Development*. 50, 2007.

Wol02. Stephen Wolfram. *A New Kind of Science*. Wolfram Media, 2002.

WW06. Matt Williams and Jon Williamson. Combining argumentation and bayesian nets for breast cancer prognosis. *Journal of Logic, Language and Information*, 2006.

Yud04. Eliezer Yudkowsky. Coherent extrapolated volition. *Singularity Institute for AI*, 2004. http://singinst.org/upload/CEV.html.

Yud06. Eliezer Yudkowsky. What is friendly ai? *Singularity Institute for AI*, 2006. http://singinst.org/ourresearch/publications/what-is-friendly-ai.html.

Zad78. L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

ZPK07. Luke S Zettlemoyer, Hanna M. Pasula, and Leslie Pack Kaelbling. Logical particle filtering. *Proceedings of the Dagstuhl Seminar on Probabilistic, Logical, and Relational Learning*, 2007.

JETGALA

LIFE BEYOND FIRST CLASS

DASSAULT'S NEW FALCON 5X

BOMBARDIER CHALLENGER 350
THE FIRST EVER GULFSTREAM
CUSTOM LIVERY DESIGN
WINGS FOR SCIENCE
NBAA 2013

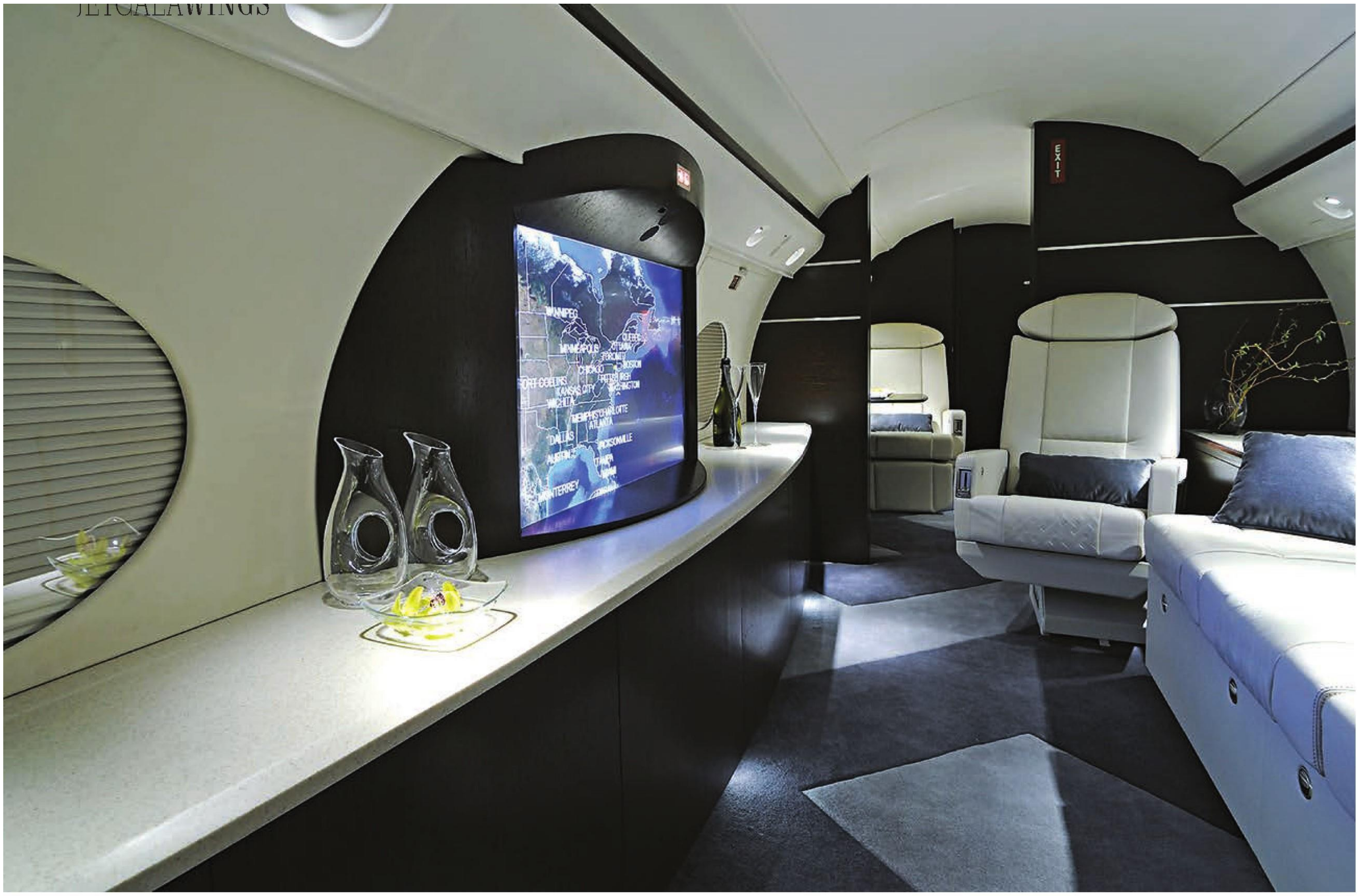
18

November 2013 – January 2014

WATCHES & WONDERS 2013

ANYA HINDMARCH | SMYTHSON BESPOKE
STOCKINGER SAFES | LOUIS MOINET
AVIATION AT HOME | SAIGON SUITES

AUD15 BND10 RMB100 HKD80 INR500 IDR85,000
KRW10,000 KWD3 MYR20 NZD20 PESOS300 QAR40
SAR40 SGD8 TWD300 BAHT250 AED40 VND100,000
8 885007 530017


EXIT
WINNIPEG
MINNEAPOLIS
OTTAWA
QUEBEC
TORONTO
BOSTON
CHICAGO
PITTSBURGH
WASHINGTON
FORT COLLINS
KANSAS CITY
WICHITA
MEMPHIS
CHARLOTTE
ATLANTA
DALLAS
AUSTIN
JACKSONVILLE
TAMPA
MONTERREY

INTERNATIONAL JET INTERIORS

by Jennifer Henricus

CABIN COUTURE

A DARING DESIGNER GOES FOR THE RUNWAY

Clients always want their aircraft to be modern and up-to-date, while maintaining an understated, luxurious aesthetic

In a 50,000-square-foot hangar at Long Island's MacArthur Airport, an aircraft interiors project has its designer "fired on all cylinders". Tasked to modify and refurbish a Global 5000 for the sports entertainment giant, World Wrestling Entertainment (WWE, formerly WWF), New York-based International Jet Interiors is putting a progressive spin on cabin design. While the livery boasts WWE's corporate colours, the interiors use "hand-woven carpets, custom-made metal finishes including a spun black pearl finish, custom-dyed leathers and bespoke seating from Italy", says Eric Roth, the designer at International Jet Interiors' helm.

But the highlight is in the high-tech facilities: studio-grade equipment that enables on-the-fly editing of feeds from live sports entertainment shows, which are transmitted via on-board satellite. Meanwhile, specially programmed iPad minis control the cabin functions. "It will be the ultimate couture craft when completed in January 2014," says Roth.

No stranger to special client requests, Roth says he loves the challenge – the more unusual the request, the more he and his team seem to excel in delivering the solution. His clients come from around the world with a varied wish-list, ranging from entertainment systems with satellite TV and high-speed WiFi to 24-karat gold-plated fixtures throughout the craft. >>

THE MORE UNUSUAL THE REQUEST, THE MORE ROTH AND HIS TEAM SEEM TO EXCEL IN DELIVERING THE SOLUTION

"HELPING CLIENTS STRATEGISE IS A KEY PART OF THE DESIGN PROCESS. I LOOK AT HOW LONG THEY HAVE OWNED THE CRAFT AND AT ITS DEPRECIATING ASSET VALUE"

Details set Roth's work apart from other aircraft interiors — from the choice of lighting, carpet, upholstery and cabinetry, to gadget stowage and dining accessories

>> Its refurbishment of Donald Trump's Boeing 757, for instance, included a 24-karat gold-plated bathroom sink. Other projects have required reconfiguring the seating layout to make kennel space for hunting dogs, installing ultra-secure bassinets for newborn infants, and wrapping a toilet seat in crocodile skin for a client in India.

Roth says that as a designer and design director, his main task is to develop an intimate understanding of a client's lifestyle and his use of the jet — the percentage of time spent for business, pleasure, corporate entertainment and family time. "Helping clients strategise is a key part of the design process. I look at how long they have owned the craft and at its depreciating asset value, and work out if it's worth it to invest in the latest piece of technology or ultra comfort accessory. Most of them will opt for the investment because they want their aircraft to be up-to-date, even if this does not up its asset value."

At times, thorough strategising is needed even before International Jet Interiors accepts a project. A US-based film producer once wanted a 'sky studio' so that he and his four-member team could edit films onboard his Gulfstream. This required state-of-the-art editing equipment, surround sound, and a 42-inch, flat-screen, high-definition TV. "The client perceived the theatre >>

SHARP
COMMERCIAL MICROWAVE OVEN

EXIT

>> aspect as the great challenge – yet it was balancing the weight of the floor-plan change and equipment furnishing that was the real challenge. The structural and technical analysis consumed the better part of three weeks before we could make the commitment to produce this unique aircraft," says Roth.

Despite many quirky refurbishing requests, Roth says a fair percentage of his clients avoid using their aircraft to make a statement: "They will come to me saying 'Eric, I want it to be luxuriously comfortable, very functional but extremely understated'." Nearly 75 per cent of the firm's clients are US-based, but it is now receiving more requests from private-jet owners in Asia. The company has worked with clients in Shanghai, Tokyo and Mumbai, and is in discussions with new customers in Africa.

He hopes to one day produce a "green and sustainable" solution for a client. That may sound like an oxymoron, and it will take a designer used to pushing the limits of convention to create such a ground-breaking design. It looks like Roth is cut out for the job.

"CLIENTS COME TO ME SAYING 'ERIC, I WANT IT TO BE LUXURIOUSLY COMFORTABLE, VERY FUNCTIONAL BUT EXTREMELY UNDERSTATED'."

OPPOSITE

International Jet Interiors achieves a sleek, streamlined appeal by modifying an aircraft's window line and periphery, and using composite materials exclusively

THIS PAGE

Hand-stitched leathers and finely woven fabrics create a luxurious, comfortable ambiance

Gadgets can be configured to control the lightings in the cabin interior

Bloomberg

Thatcher, Mandela, Chavez Are Among Notable Deaths in 2013

2013-12-26 05:01:00.3 GMT

By Steven Gittelson

Dec. 26 (Bloomberg) -- The first female prime minister of the U.K., the first black president of South Africa and the first woman to buy a seat on the New York Stock Exchange were among the notable deaths of 2013.

Margaret Thatcher, 87, died in April; Nelson Mandela, 95, died this month; and Muriel Siebert, 84, died in August.

The year also included the deaths of politicians Edward Koch, 88, in February and Hugo Chavez, 58, in March; musicians Marian McPartland, 95, in August and Lou Reed, 71, in October; and athletes Stan Musial, 92, in January, and Ken Norton, 70, in September.

The world of business, finance and investing lost Fred Turner, 80, the former McDonald's Corp. chief executive officer who introduced Chicken McNuggets, Egg McMuffins and Happy Meals, in January; Martin Zweig, 70, who predicted the 1987 stock-market crash, in February; and Alfred Feld, 98, whose 80 years at Goldman Sachs Group Inc. made him the firm's longest-serving employee, in November.

Here are the year's notable deaths, with each name linked to a previously published obituary. A cause of death is provided when known.

January

Patti Page, 85. U.S. pop singer whose 1950s hits included "Tennessee Waltz" and "(How Much Is That) Doggie in the Window?" Died Jan. 1.

Lewis Adam, 68. He was a fuel trader who became president of ADMO Energy LLC, a supply consultant in Kansas City, Missouri.

Died Jan. 2 of a heart attack on his first day of retirement.

Fred Turner, 80. As CEO at McDonald's Corp., now the world's largest restaurant company, he introduced Chicken McNuggets, the Egg McMuffin and Happy Meals. Died Jan. 7 of complications from pneumonia.

Ada Louise Huxtable, 91. She became the first full-time architecture critic at a U.S. newspaper when she was hired by the New York Times in 1963 and won the first Pulitzer Prize for distinguished criticism, in 1970. Died Jan. 7.

James M. Buchanan, 93. The U.S. economist who won the 1986 Nobel
Prize for applying the tools of economics to analyze political decision-making. Died Jan. 9.

Daniel J. Edelman, 92. He founded Chicago-based Daniel J. Edelman Inc., now the world's largest independent public-relations company, and helped pioneer the use of celebrities in PR campaigns. Died Jan. 15 of heart failure.

Thomas Candillier, 37. The Paris-based head of European equity sales at JPMorgan Chase & Co., who joined the bank in 2001 after working in energy derivative sales at Goldman Sachs. Died Jan. 16.

Robert Citron, 87. He was the treasurer of Orange County, California, in 1994, when his bad bets on derivative securities lost about $1.7 billion, causing what was then the biggest U.S. municipal bankruptcy. Died Jan. 16.

Pauline Phillips, 94. To millions of U.S. newspaper readers, she was Abigail Van Buren, author of the personal advice column, "Dear Abby." Died Jan. 16 from Alzheimer's disease.

Stan Musial, 92. A Hall of Fame outfielder for Major League Baseball's St. Louis Cardinals, "Stan the Man" was one of the game's great hitters during the 1940s and 1950s. Died Jan. 19.

Earl Weaver, 82. He was the hot-tempered manager of the Baltimore Orioles baseball team for 17 years, guiding his club to the World Series four times and winning the championship in 1970. Died Jan. 19 of a heart attack.

Michael Winner, 77. The British film director best known for making the first three "Death Wish" action movies, starring Charles Bronson. Died Jan. 21 of liver cancer.

A.W. "Tom" Clausen, 89. He rose from part-time cash counter to CEO at Bank of America Corp., and returned for a second stint as chief after serving as World Bank president. Died Jan. 21 of complications from pneumonia.

Maria Schaumayer, 82. The Austrian economist who in 1990 became the first woman to lead a European central bank. Died Jan. 23.

John M. "Jack" McCarthy, 85. The stock-market optimist who from 1983 to 1992 was co-managing partner at Lord Abbett & Co., an investment management firm, in Jersey City, New Jersey. Died Jan. 23.

Barry Lind, 74. Founder of Lind-Waldock & Co., a discount futures firm in Chicago, who helped transform the Chicago Mercantile Exchange into a market for financial futures. Died Jan. 24, one day after he was struck by a car.

Ben Steele, 35. He joined London-based hedge fund Armajaro Asset Management LLP in 2012 to start a pool trading shares of financial companies. Died Jan. 25 of an apparent heart attack.

Stefan Kudelski, 84. The Polish-born inventor of the first professional-quality portable audio recorder, in 1951. Died Jan. 26 in Switzerland.

Patty Andrews, 94. Last surviving member of the Andrews Sisters trio, the most popular female vocal group of the first half of the 20th century. Died Jan. 30 in Los Angeles.

Caleb Moore, 25. A Texas-born snowmobile racer who became the sport’s first fatality. Died Jan. 31, one week after crashing at the Winter X Games in Aspen, Colorado.

February

Edward I. Koch, 88. As New York mayor from 1978 to 1989, he led the city back from the brink of bankruptcy, turning a $1 billion budget deficit into a $500 million surplus in five years. Died Feb. 1 of heart failure.

Edith Lauterbach, 91. Last survivor of a quintet of U.S. women, who in 1945 founded the Air Line Stewardesses Association, the world’s first union for flight attendants. Died Feb. 4.

George Frazer, 86. Chairman of Toronto-based Leon Frazer & Associates, who invested in companies with high dividends during his seven decades as a fund manager. Died Feb. 6.

Rem Vyakhirev, 78. He was CEO of OAO Gazprom, the world’s biggest natural-gas producer, from 1993 to 2001. Died Feb. 11.

Stokley Towles, 77. He spent his career at the Boston office of New York-based Brown Brothers Harriman & Co., creating the firm’s global custody service, which now accounts for more than 70 percent of the bank’s employees. Died Feb. 14.

Mindy McCready, 37. A U.S. country music singer whose hits included “Guys Do It All the Time.” Died Feb. 17 of apparent suicide.

Otto Beisheim, 89. German billionaire, who in 1964 co-founded Metro AG, now Germany’s biggest retailer. Died Feb. 18 of suicide after suffering from an incurable illness.

Jerry Buss, 80. After buying the Los Angeles Lakers in 1979, he added marquee stars including Magic Johnson and Kobe Bryant, winning 10 National Basketball Association championships between 1980 and 2010. Died Feb. 18 of kidney failure related to cancer treatment.

Martin Zweig, 70. He predicted the 1987 stock-market crash and wrote books and newsletters that influenced U.S. investors for more than a quarter century. Died Feb. 18.

Alger “Duke” Chapman Jr., 81. The CEO of Shearson Hammill & Co., who merged the firm with Sanford Weill’s Hayden Stone Inc. in 1974, a milestone in the emergence of mega-companies within the finance industry. Died Feb. 18 of congestive heart failure in Little Rock, Arkansas, where he retired.

Susan Carroll, 50. A managing director of Morgan Stanley since 2009, who was the chief operating officer of Salt Lake City-based Morgan Stanley Bank. Died Feb. 18 of liver disease.

Paul McIlhenny, 68. He was the fourth generation of his family to lead McIlhenny Co., a maker of Tabasco sauce. Died Feb. 23 of a heart attack at his home in New Orleans.

C. Everett Koop, 96. As U.S. surgeon general from 1981 to 1989, he used his position to educate Americans about the dangers of smoking while pushing the government to take a stronger stand against AIDS. Died Feb. 25.

Stephane Hessel, 95. A hero of the French Resistance and former United Nations diplomat, who in 2010 wrote “Indignez-Vous!,” titled “Time for Outrage” in the U.S., a best-selling pamphlet that helped inspire social protests in Europe and the Occupy Wall Street movement. Died Feb. 26 in Paris.

Robert Elberson, 84. As CEO of Hanes Corp., he introduced L’eggs pantyhose, and became president of Hanes’s parent, Sara Lee Corp. Died Feb. 26 at his home in Charlotte, North Carolina.

Van Cliburn, 78. The pianist from Texas, whose triumph as a 23-year-old at the 1958 Tchaikovsky International Piano and Violin Festival in Moscow made him an international star. Died Feb. 27 of bone cancer.

Bruce Reynolds, 81. Mastermind of the 1963 Great Train Robbery in Britain, which brought him fame, fortune and 10 years in prison. Died Feb. 28.

March

Bonnie Franklin, 69. The actress best known for playing divorced mother Ann Romano in the U.S. television show “One Day at a Time,” which aired from 1975 to 1984. Died March 1 of pancreatic cancer.

James Strong, 68. Former CEO of Qantas Airways Ltd., Australia’s biggest airline, and former chairman of Woolworths Ltd., the country’s largest retailer. Died March 3 in Sydney of complications from surgery.

Hugo Chavez, 58. President of Venezuela since 1998, who used the country’s oil wealth to help the poor, nationalized corporations and dismissed foes as puppets of U.S. imperialism. Died March 5 of cancer.

John J. Byrne, 80. He led Geico Corp. from 1976 to 1985 and saved the insurer from bankruptcy, leading Warren Buffett to buy the company and call him “the Babe Ruth of insurance.” Died March 7 of prostate cancer.

Michael P. Duffy, 54. As the head of JPMorgan’s Dallas-based Chase Paymentech unit, he helped make the bank one of the largest U.S. processors of credit-card and electronic payments. Died March 7 of cancer.

Ewald-Heinrich von Kleist, 90. Last survivor of a group of German army officers, who tried unsuccessfully to kill Adolf Hitler. Died March 8.

Elizabeth Cheval, 56. She founded EMC Capital Management Inc., a Bannockburn, Illinois-based investment firm. Died March 9 in China after suffering a brain aneurysm during a business trip.

Ieng Sary, 87. Former foreign minister of the Khmer Rouge, who died while on trial for the deaths of 1.7 million Cambodians in the 1970s. Died March 14.

K. Anji Reddy, 74. Billionaire founder of Dr. Reddy's Laboratories Ltd., India's second-largest drugmaker. Died March 15 of liver cancer.

Olivier Metzner, 63. One of France's best-known defense lawyers, whose clients included former Societe Generale SA trader Jerome Kerviel. Died of suicide on March 17, when his body was found floating near his private island in Brittany.

Steve Davis, 60. Quarterback on the University of Oklahoma's national college football championship teams in 1974 and 1975. Died March 17 in a plane crash.

Harry Reems, 65. The male star of "Deep Throat," a 1972 U.S. film that brought hardcore pornography to mainstream audiences. Died March 19 in Salt Lake City.

Rise Stevens, 99. New York City-born mezzo-soprano who starred at the Metropolitan Opera in the 1940s and 1950s and was best known for playing the lead role in "Carmen." Died March 20.

Chinua Achebe, 82. The Nigerian author of "Things Fall Apart" (1958), one of the first novels by an African writer to attract a worldwide audience. Died March 21.

Georg W. Claussen, 100. The CEO at Beiersdorf AG, the Hamburg-based maker of Nivea skin cream, from 1957 to 1979. Died March 21.

Ray Williams, 58. A former guard in the National Basketball Association whose 10-year career included stints with the New York Knicks and New Jersey Nets. Died March 22.

Boris Berezovsky, 67. He was one of the first and best-known oligarchs who accumulated vast wealth and influence in post-Soviet Russia until a falling out with Russian President Vladimir Putin. Died March 23 at his home near London, where he lived in self-imposed exile.

Virgil “Fire” Trucks, 95. Hurled two no-hitters in 1952 for Major League Baseball’s Detroit Tigers. Died March 23.

Anthony Lewis, 85. Former New York Times reporter and columnist, who won two Pulitzer Prizes and transformed coverage of the U.S. Supreme Court. Died March 25 of renal and heart failure.

Guillermo Luksic Craig, 57. Chairman of Chilean holding company Quinenco SA, and a member of Chile’s richest family. Died March 27 of lung cancer.

Ralph Klein, 70. The premier of Alberta, Canada’s oil-rich province, from 1992 to 2006. Died March 29 of dementia and lung disease.

Mal Moore, 73. He was part of 10 national championship college football teams as a player, coach and athletic director at the University of Alabama. Died March 30 of pulmonary disease.

April

Jack Pardee, 76. The All-American linebacker at Texas A&M University, who played in the NFL and then coached the league’s Chicago Bears, Washington Redskins and Houston Oilers. Died April 1 of gall bladder cancer.

Barbara Piasecka Johnson, 76. The Polish-born cook and chambermaid who married Johnson & Johnson heir J. Seward Johnson, won $350 million in a legal battle with his children over his will, and wound up a billionaire art collector and philanthropist living in Monaco. Died April 1 in Poland.

Calvert Crary, 69. A Wall Street lawyer whose newsletter, “Litigation Notes,” predicted the outcome of corporate court battles for an audience of hedge fund managers and institutional investors. Died April 6.

Margaret Thatcher, 87. The U.K. prime minister from 1979 to 1990, known as the “Iron Lady” for her strong will, who helped end the Cold War and revived Britain’s economy by deregulating financial markets, lowering taxes and privatizing companies. Died April 8 of a stroke in London.

Annette Funicello, 70. She was the most popular of the original Mouseketeers on Walt Disney's "The Mickey Mouse Club" television show in the 1950s, then had a career as an actress and singer. Died April 8 of complications from multiple sclerosis, in California.

Robert G. Edwards, 87. A British physiologist, his research on in-vitro fertilization led to the first test-tube baby, earning him the Nobel Prize for Medicine in 2010. Died April 10.

George Schaefer, 84. As chairman and CEO of Caterpillar Inc. from 1985 to 1990, he led the construction equipment maker from losses to record profits. Died April 10 in Peoria, Illinois, the company's hometown.

Jonathan Winters, 87. The American stand-up comic, whose improvisational humor, starting in the 1950s, inspired comedians such as Robin Williams and Jim Carrey. Died April 11.

Maria Tallchief, 88. One of the premier U.S. ballerinas of the 20th century, and the wife of choreographer George Balanchine. Died April 11.

Colin Davis, 85. The British-born principal conductor of the London Symphony Orchestra from 1995 to 2006. Died April 14.

George Beverly Shea, 104. Known as "America's Beloved Gospel Singer," he performed before more than 200 million people during six decades with evangelist Billy Graham. Died April 16.

Pat Summerall, 82. The former NFL player, who teamed with John Madden for 21 years to form one of the most popular broadcasting pairings in television history. Died April 16.

Al Neuharth, 89. He built Gannett Co. into the largest U.S. newspaper publisher and created USA Today, which became the country's biggest-selling daily paper. Died April 19 of complications from a fall.

Cortright McMeel, 41. He drew on his experience in the commodities market to write a darkly comic novel about energy traders. Died April 19 in Denver, where he lived.

Dirce Navarro de Camargo, 100. She became Brazil's richest woman after inheriting Camargo Correa SA, now the nation's third-largest construction company, founded by her late husband, Sebastiao Camargo. Died April 20.

Richie Havens, 72. The Brooklyn-born folk singer best known as the opening act at the Woodstock music festival in 1969. Died April 22 of a heart attack.

Kathryn Wasserman Davis, 106. She gave her husband about $100,000 in 1947 to open his own investment firm, Shelby Cullom Davis & Co., which was valued at $800 million when he died in 1994. Died April 23.

George Jones, 81. The country-music singer, whose emotion-drenched vocal style earned him more hit records than any other artist. Died April 26 in Nashville, Tennessee.

Tim Taylor, 71. He was Yale University's ice hockey coach for 28 seasons, winning six Ivy League titles. Died April 27 of cancer.

Janos Starker, 88. A Hungarian-born child prodigy, who became one of the most renowned cellists of the 20th century and ended his career as a distinguished professor of music at Indiana University. Died April 28.

Edward Feigeles, 58. He was a managing director at Lehman Brothers Holdings Inc. in New York, where he led the private client services group, from 1996 to 2005. Died April 29 after a brief illness.

Bill Mahoney, 55. He led global sales and marketing at Westport, Connecticut-based Bridgewater Associates LP, the world's largest hedge fund, before leaving in 2006. Died April 30 of pancreatic cancer.

May

William Cox Jr., 82. The patriarch of the Bancroft clan that for 105 years controlled New York-based Dow Jones & Co., publisher of the Wall Street Journal, who helped persuade the extended family to sell the company to Rupert Murdoch's News Corp. in 2007. Died on May 1 of complications from diabetes.

Giulio Andreotti, 94. The seven-time Italian prime minister, whose political career embodied the highs and lows of Italian postwar governance. Died May 6.

Salvatore J. Trani, 72. He helped rebuild the credit unit of Cantor Fitzgerald LP's BGC Partners Inc. after 658 of the firm's employees were killed in the Sept. 11, 2001, attacks on New York's World Trade Center. Died May 7 of brain cancer.

George Sauer Jr., 69. Playing wide receiver for the National Football League's New York Jets, he caught eight passes from quarterback Joe Namath in the 1969 Super Bowl, which the underdog Jets won. Died May 7 of Alzheimer's disease.

Alan Abelson, 87. A U.S. financial journalist, who was a former top editor at Barron's magazine and wrote a widely followed stock-market column. Died May 9 of a heart attack.

Andrew Simpson, 36. A British sailor who won two Olympic medals in sailing for Britian. Died May 9 when a yacht attempting to compete for the America's Cup capsized in San Francisco Bay.
Ottavio Missoni, 92. The founder of Missoni SpA, a high-end Italian fashion company. Died May 9.

Deborah Bernstein, 41. She became a partner at Aquiline Capital Partners LLC, a New York-based private equity firm, after starting her career at Goldman Sachs. Died May 10 of cancer.

Walter J. O'Brien III, 46. The head of equities sales and trading at BB&T Corp., North Carolina's second-biggest bank, and a mentor to finance-minded graduates of the University of Richmond, his alma mater. Died May 10 of colon cancer.

Joyce Brothers, 85. Armed with a Ph.D. in psychology, she became a pioneer in dispensing advice about love, self-image and sex on U.S. television and radio and syndicated newspaper columns, starting in the late 1950s. Died May 13 of respiratory failure.

Chuck Muncie, 60. A running back for the NFL's New Orleans Saints and San Diego Chargers who in 1981 set a record for rushing touchdowns in a season and then had his career cut short because of cocaine use. Died May 13 of a heart attack.

Jorge Rafael Videla, 87. Argentina's military junta leader, who oversaw a campaign of murder and kidnapping from 1976 to 1981. Died May 17 in a Buenos Aires prison, where he was serving a life sentence for human rights violations.

Ken Venturi, 82. The American golfer who won the 1964 U.S. Open and spent 35 years as a TV golf analyst. Died May 17 of a spinal infection and pneumonia.

Isabel Benham, 103. Her mastery of U.S. railroad financing in the 1930s made her an influential bond analyst and in 1964 she became the first female partner at R.W. Pressprich & Co., a Wall Street firm. Died May 18.

Ray Manzarek, 74. The keyboardist and songwriter who with Jim Morrison founded The Doors, a 1960s U.S. rock group that sold more than 100 million records. Died May 20 of bile duct cancer.

John Q. Hammons, 94. He created John Q. Hammons Hotels in 1969 and became the largest private, independent upscale-hotel manager in the U.S., developing properties for brands such as Marriott, Renaissance and Embassy Suites. Died May 26.

Roberto Civita, 76. An Italian-born entrepreneur, who became the billionaire chairman of Grupo Abril, which publishes some of Brazil's most-read magazines. Died May 26 of complications from an abdominal aneurysm.

Cullen Finnerty, 30. The former college football star who led Grand Valley State University's team to three Division II national championships and won more than 50 games over four seasons as the school's quarterback. Died between May 26 and May 28 of pneumonia after disappearing while on a fishing trip in Michigan.

Charles Henderson, 88. He was the fourth generation of his family to run Henderson Brothers Inc., a specialist floor trading firm on Wall Street. Died May 29 of heart failure.

George H. Weiler III, 69. Senior vice president for wealth-management services at UBS AG, who started his Wall Street career in 1984 at Dillon Read & Co. Died May 29 of a heart attack.

June

Raymond Saxe, 50. A former senior vice president of global risk technology at HSBC Holdings Plc, who worked on Wall Street for 21 years. Died June 1.

Michael McClintock, 55. Senior managing director in New York for Macquarie Group Ltd., Australia's biggest investment bank. Died June 2 of cardiac arrest.

Chen Xitong, 82. The mayor of Beijing during the 1989 Tiananmen Square protests in which hundreds of people were killed. Died June 2 of cancer.

Hugh P. Lowenstein, 82. A managing director of Donaldson, Lufkin & Jenrette in the 1990s, founder of Shore Capital Ltd. in Bermuda and a director of Bloomberg LP, parent company of Bloomberg News, for more than 15 years. Died June 2.

Frank Lautenberg, 89. The five-term Democratic senator from New Jersey who wrote laws raising the legal drinking age to 21 and banning smoking on domestic airline flights. Died June 3 of complications from viral pneumonia.

Deacon Jones, 74. A Hall of Fame defensive end, who was the NFL's defensive player of the year in 1967 and 1968 when he played for the Los Angeles Rams. Died June 3.

Esther Williams, 91. The U.S. swimming champion who was best known as a movie actress in aquatic musicals in the 1940s and 1950s. Died June 6.

William L. Clayton, 83. During his 55-year career on Wall Street, he spent almost four decades at E.F. Hutton & Co. and founded Hutton Capital Management. Died June 7 of Parkinson's disease.

Robert Fogel, 86. The University of Chicago economist, who won a Nobel Prize in 1993 for his historical analysis of how railroads and slavery shaped U.S. economic history. Died June 11.

Miller Barber, 82. A U.S. golfer who made a record 1,297 combined starts on the U.S. PGA and Champions golf tours, winning 35 titles. Died June 11.

Jiroemon Kimura, 116. He was recognized by Guinness World Records as the oldest male in recorded history. Died June 12 in his hometown of Kyotango, in western Japan.

Paul Soros, 87. The Hungarian-born founder of Soros Associates, a New York-based builder of shipping ports, and the older brother of billionaire investor George Soros. Died June 15.

Mathew Gladstein, 90. Working with future Nobel winners Robert Merton and Myron Scholes, he helped popularize options trading while working at Donaldson Lufkin & Jenrette in New York. Died June 18.

Gyula Horn, 80. The prime minister of Hungary from 1994 to 1998, who as foreign minister in 1989 helped trigger events that led to the fall of the Berlin Wall. Died June 19.

James Gandolfini, 51. The New Jersey-born actor best known for portraying the conflicted mob boss Tony Soprano in the TV series "The Sopranos." Died June 19 of a heart attack while on vacation in Rome.

Dave Jennings, 61. An All-Pro punter who played for New York's two NFL teams, the Giants (1974 to 1984) and Jets (1985 to 1987). Died June 19 from Parkinson's disease.

Allan Simonsen, 34. A Danish race car driver affiliated with the Aston Martin Racing Team. Died June 22 when his car crashed at the Le Mans 24 hours race.

Bobby "Blue" Bland, 83. A Tennessee-born singer of Southern blues and ballads in hit singles such as "Turn on Your Love Light," who was inducted into the Rock and Roll Hall of Fame. Died June 23.

Harry Parker, 77. He coached the Harvard University men's heavyweight crew team to 22 undefeated seasons and eight national titles. Died June 25 of a blood disorder.

Marc Rich, 78. The Belgium-born commodities trader, who in 1983 was indicted for U.S. income tax evasion and racketeering, fled the country and lived as a fugitive until pardoned by President Bill Clinton in 2001. Died June 26 near his home in Switzerland.

Rawleigh Warner, 92. As the chairman and CEO of Mobil Oil Corp. from 1969 to 1986, he outmaneuvered competitors to make Mobil second in sales behind Exxon Corp., years before the companies merged. Died June 26 of complications from a progressive muscle disease.

July

William H. Gray III, 71. He was a Democrat from Philadelphia who served six terms in the U.S. House of Representatives, becoming the first black party whip, the No. 3 leadership position. Died July 1 while in London to attend the Wimbledon tennis tournament.

Andrew McMenigall, 47. A senior global equities manager at Scotland’s Aberdeen Asset Management Plc, who was based in Edinburgh. Died July 2 in a traffic accident while participating in a charity bicycle ride across Britain.

Toby Wallace, 36. He was a Philadelphia-based senior relationship manager at Aberdeen Asset Management Plc. Died July 2 of injuries suffered from a traffic accident while taking part in a charity bicycle ride across Britain.

Douglas Engelbart, 88. The U.S. electrical engineer who invented the computer mouse, the design of which was described in a patent filed in 1967 and granted in 1970. Died July 2 of kidney failure.

Cynthia Lufkin, 51. The philanthropist wife of Dan Lufkin, co-founder of Wall Street firm Donaldson, Lufkin & Jenrette. Died July 3 of complications from breast and lung cancer.

Douglas J. Dayton, 88. The son of a successful U.S. retailer, he became the first president of Target, a U.S. department store chain. Died July 5 of cancer.

Masao Yoshida, 58. The plant manager of Japan’s Fukushima nuclear reactor in March 2011, when an earthquake and ensuing tsunami crippled the facility in the worst nuclear disaster since Chernobyl in eastern Europe. Died July 9 of esophageal cancer.

Philip Caldwell, 93. He was the first CEO of Ford Motor Co. who wasn’t a member of the Ford family. Died July 10 of complications from a stroke.

Amar Bose, 83. An engineer who taught at Massachusetts Institute of Technology for more than four decades, he was best known as the billionaire founder of Bose Corp., an audio technology company specializing in speakers and headphones located in Framingham, Massachusetts. Died July 12.

Cory Monteith, 31. The Canadian-born actor was best known for starring in the hit TV show “Glee.” Died July 13 of a drug overdose.

Herbert Allison Jr., 69. He was the former president of Merrill Lynch & Co., chairman and CEO of TIAA-CREF, CEO of Fannie Mae and led the U.S. government’s bank bailout program. Died July 14.

Neal McCabe, 60. He was a Boston-born former global co-head of a Lehman Brothers Holdings Inc. unit focused on increasing trades with security dealers worldwide. Died July 17, two months after suffering a stroke.

Donald J. Mulvihill, 56. He was a managing director at Goldman Sachs, who started the firm’s asset-management business in Japan and created tax-focused funds in the U.S. during his 33-year career with the bank. Died July 19 of leukemia in Illinois, where he was born and raised.

Helen Thomas, 92. The pioneering female journalist who worked as White House correspondent for United Press International, where she worked for 57 years, and as a columnist for Hearst Newspapers. Died July 20.

Carsten Schloter, 49. The German-born CEO of Swisscom AG, Switzerland’s biggest telecommunications company, since 2006. Died July 23 of what police called an apparent suicide.

Emile Griffith, 75. Former U.S. welterweight and middleweight boxing champion best known for his fatal knockout of Benny Paret in a nationally televised fight in 1962. Died July 23.

Dennis Dammerman, 67. He was CEO Jack Welch’s right-hand man at General Electric Co., where at age 38 he became the company’s youngest chief financial officer and then ran GE Capital. Died July 23.

Arthur Makadon, 70. The chairman of Ballard Spahr LLP, a Philadelphia-based law firm, from 2002 to 2011. Died July 24 of cancer.

Virginia Johnson, 88. One of the key figures in the sexual revolution in postwar America, she conducted groundbreaking research in human sexuality with her collaborator, William Masters. Died July 24.

Barnaby Jack, 36. A New Zealand-born computer-security professional who exposed how hackers could attack bank automated teller machines, insulin pumps and other electronic devices. Died July 25.

Lindy Boggs, 97. She spent 18 years in the U.S. House of Representatives, succeeding her husband, Hale Boggs, and worked as a champion for women's rights. Died July 27.

George "Boomer" Scott, 69. Large, strong and agile, he spent nine of his 14 seasons in Major League Baseball with the Boston Red Sox, playing first base and leading the team to win the American League pennant in 1967. Died July 28.

Peter Flanigan, 90. The former Dillon Read investment banker, who worked as deputy campaign manager for Richard Nixon's successful 1968 presidential run, then joined the administration as an adviser on business and economic matters. Died July 29.

Berthold Beitz, 99. German industrialist who hid Jews from the Nazis during World War II and then helped rebuild Fried Krupp GmbH, a predecessor of the country's biggest steelmaker. Died July 30.

August

Art Donovan, 88. An NFL Hall of Fame defensive tackle who won two championships with the Baltimore Colts in the 1950s. Died Aug. 4.

E. Nelson Asiel, 96. Third-generation leader of Asiel & Co., a Wall Street brokerage firm founded by his grandfather, Elias, in 1878. Died Aug. 5.

Jerry Wolman, 86. He owned the NFL's Philadelphia Eagles team from 1963 to 1969. Died Aug. 6.

Karen Black, 74. The U.S. actress best known for her performances in "Five Easy Pieces," "Easy Rider" and "Nashville." Died Aug. 8 of cancer.

Lorraine Lodge, 52. She was a convertible bond specialist during a career in New York and London at Merrill Lynch, ING Barings and Nomura Holdings Inc. Died Aug. 8 of ovarian cancer in New York, where she lived.

Lee Quo-wei, 95. The former chairman of Hong Kong's Hang Seng Bank Ltd. and in 1969 was part of the group that created the Hang Seng Index, the city's benchmark stock gauge. Died Aug. 10.

Eydie Gorme, 84. American pop music singer best known for her 1963 hit "Blame It on the Bossa Nova," and for nightclub and television performances with her husband, Steve Lawrence. Died Aug. 10.

Friso van Oranje, 44. A member of the royal family in the Netherlands, he gave up his place in line for the throne to marry the woman he loved. Died Aug. 12 of complications from brain damage suffered in a skiing accident in February 2012.

John H. Laporte, 68. He worked at Baltimore-based T. Rowe Price Group Inc. from 1976 until retiring in 2012 and was named mutual fund manager of the year in 1995. Died Aug. 12 of complications from lymphoma.

Louis V. Gerstner III, 41. The son of Louis Gerstner Jr., the former CEO of International Business Machine Corp., and president of the Gerstner Family Foundation. Died Aug. 14 after choking while dining in a New York restaurant.

Elmore Leonard, 87. Known as the "Dickens of Detroit," Leonard was the best-selling author of crime novels and Westerns, many of which were made into movies, including "Get Shorty" and

“Hombre.” Died Aug. 20 of complications from a stroke.

Marian McPartland, 95. The British-born jazz pianist, whose National Public Radio show, in which she interviewed and played with musicians from Benny Goodman to Elvis Costello, was broadcast for more than three decades. Died Aug. 20 at her home in New York.

Ronald L. Motley, 68. A South Carolina lawyer, he led lawsuits against tobacco companies, resulting in a payout of $246 billion, the biggest civil settlement in U.S. history. Died Aug. 22 from complications of organ failure.

Julie Harris, 87. The U.S. actress who appeared in 30 Broadway plays and won five Tony awards. Died Aug. 24 of congestive heart failure.

Muriel Siebert, 84. The first woman to buy a seat on the New York Stock Exchange, in 1967, founder of Muriel Siebert & Co., a discount brokerage, and the first female superintendent of banks for New York State. Died Aug. 24 of complications from cancer.

Eric T. Miller, 85. The former chief investment officer for New York-based Donaldson Lufkin & Jenrette, who called the stock-market bottom in 1982 and whose “Random Gleanings ” market commentary was widely followed by investors. Died on Aug. 29 of complications from brain cancer.

Seamus Heaney, 74. Irish poet who won the Nobel Prize in Literature in 1995. Died Aug. 30.

David Brenneman, 37. He was an executive director in equity risk management at Morgan Stanley in New York, who previously worked at Banc of America Securities and Davis Polk & Wardwell LLP. Died Aug. 31 of cancer.

David Frost, 74. The British television interviewer best known for his 1977 interviews with former President Richard Nixon, which became the basis for the 2008 movie “Frost/Nixon.” Died Aug. 31 of a heart attack aboard the Queen Elizabeth cruise ship.

September

Tommy Morrison, 44. In 1993, he defeated George Foreman to win the World Boxing Organization heavyweight title and appeared in the movie “Rocky V.” Died Sept. 1.

Ronald Coase, 102. The British-born University of Chicago economist who won the Nobel Prize in 1991 for research he said showed that “people will use resources in the way that produces the most value.” Died Sept. 2.

Joseph Granville, 90. He was a U.S. financial newsletter writer and technical analyst who moved stock markets with bearish calls in the 1970s and 1980s. Died Sept. 7.

Ray Dolby, 80. He was a U.S. inventor who became a billionaire by designing noise-reduction and surround-sound technologies used in films, movie theaters and home-theater equipment. Died Sept. 12 of leukemia.

Ken Norton, 70. The U.S. boxer who was a former world heavyweight champion and gained fame by breaking Muhammad Ali’s jaw during a match. Died Sept. 18 after suffering a series of strokes.

Joy Covey, 50. She joined Amazon.com Inc. during its pioneering days as an Internet retailer, serving as its chief financial officer when the company held its initial public offering in 1997. Died Sept. 18 in a bicycle accident in California, where she lived.

Hiroshi Yamauchi, 85. The great-grandson of Nintendo Co.’s founder, running the company for 53 years and becoming Japan’s richest person in 2008. Died Sept. 19.

Douglas Millett, 49. He was director of research at New York-based Kynikos Associates Ltd., who called Enron Corp. “a hedge fund sitting on top of a pipeline” and helped expose the energy company’s financial problems. Died Sept. 21 of cancer.

Richard T. McSherry, 77. The co-founder, along with James Elkins, of New York-based Elkins/McSherry LLC, which pioneered a way to crunch data to assess trading costs and help institutional investors maximize profits. Died Sept. 26 of prostate cancer.

L.C. Greenwood, 67. The four-time NFL Super Bowl champion, who played defensive end on the Pittsburgh Steelers’ defensive line known as the “steel curtain” in the 1970s. Died Sept. 29.

October

Tom Clancy, 66. The U.S. author of “The Hunt for Red October” and “Patriot Games,” he became one of the world’s best-known writers by infusing espionage thrillers with technical details about military weaponry and intelligence agencies. Died Oct. 1.

Karen Strauss Cook, 61. In 1975, she became the first woman hired in Goldman Sachs’s equities division, and the firm’s first female trader. Died Oct. 2 of a degenerative brain disease in New York, where she lived.

Amy Dombroski, 26. The U.S. bicyclist who was a three-time national cyclo-cross champion. Died Oct. 3 when struck by a vehicle while training in Belgium.

Sergei Belov, 69. He played guard on the Soviet Union’s basketball team that beat the U.S. to win a gold medal in the 1972 Olympics. Died Oct. 3.

Vo Nguyen Giap, 102. The North Vietnamese general whose fighters drove the French out of Vietnam in 1954, then served as commander-in-chief against U.S. forces during the Vietnam War. Died Oct. 4.

Ovadia Yosef, 93. An ultra-Orthodox rabbi who galvanized Israel’s Jews of Middle Eastern and North African descent into a political force with the Shas Party. Died Oct. 7.

Paul Desmarais Sr., 86. The Canadian billionaire, who turned an inherited fleet of buses into Power Corp. of Canada, an insurance and financial services conglomerate. Died Oct. 8.

Scott Carpenter, 88. The second American to orbit the Earth, he was one of the original seven astronauts in Project Mercury, the first U.S. human spaceflight program. Died Oct. 10 of complications from a stroke.

Wilfried Martens, 77. A former prime minister of Belgium, who presided over nine governments from 1979 to 1992, deepening the nation's integration in the European Union while leaving a legacy of debt. Died Oct. 10.

Wally Bell, 48. A Major League Baseball umpire for 21 years. Died Oct. 14 of a heart attack.

Hans Riegel, 90. The German billionaire owner of Haribo GmbH, a candy maker started by his father, whose best-known product is the Gummy Bear. Died Oct. 15 of heart failure.

Peter A. Levy, 77. He followed the path of his father, Gustave Levy, becoming a partner at Goldman Sachs, until departing to co-found investment funds, including Harmony Capital Management LP, a New York-based fund of private-equity funds. Died Oct. 18 of cancer.

Tom Foley, 84. He was a Democratic congressman from Washington State from 1964 to 1994 and rose to speaker of the House. Died Oct. 18 of pneumonia following a series of strokes.

Sally Dawson, 39. A British-born banker who spent 17 years at the London office of Deutsche Bank AG, specializing in high-yield and distressed-debt sales. Died Oct. 18 of cancer.

C.W. “Bill” Young, 82. A U.S. representative from Florida, he was the longest-serving Republican in Congress and an advocate of military spending. Died Oct. 18 of complications following surgery.

Oail “Bum” Phillips, 90. A Texan who spent 12 seasons as a coach in the NFL for the Houston Oilers and New Orleans Saints, pacing the sidelines in cowboy boots, jeans and a Stetson hat. Died Oct. 18.

William C. Lowe, 72. He supervised the production of International Business Machines Corp.’s first personal computer, in 1980. Died Oct. 19 of a heart attack.

Lawrence Klein, 93. The U.S. economist who won the 1980 Nobel Prize for developing computer models to help predict global economic trends. Died Oct. 20.

Jamalul Kiram III, 75. A Philippine sultan, who waged an armed struggle for control over Malaysia’s Sabah state, an area rich in natural resources. Died Oct. 20 of kidney disease.

Juliette Moran, 96. She joined GAF Corp. in 1943 when it was a New York-based chemical maker, rising to vice chairman in 1980. Died Oct. 20.

Don James, 80. In 1975, he became the head coach of the University of Washington’s football team, winning a share of the national title in 1991. Died Oct. 20 of pancreatic cancer.

K.S. “Bud” Adams, 90. Owner of the NFL’s Houston Oilers team and its successor, the Tennessee Titans, he helped found the American Football League in 1960. Died Oct. 21.

Anthony Caro, 89. A British sculptor, who created large art objects with heavy steel girders, metal sheets, pipes and scrap metal and was knighted in 1987. Died Oct. 23 of a heart attack.

Paul Reichmann, 83. One of three brothers who built Toronto-based Olympia & York Developments Ltd. in building London’s Canary Wharf and New York’s World Financial Center before it filed for bankruptcy in 1992. Died Oct. 25.

Bill Sharman, 87. He was elected to the Basketball Hall of Fame twice, first as a player, in 1976, and then as a coach, in 2004, a feat achieved only by John Wooden and Lenny Wilkens. Died Oct. 25 following a stroke.

Kimberly Mounts, 48. She founded MAP Alternative Asset Management Co. in Newport Beach, California, in 2006, following jobs at Goldman Sachs and Morgan Stanley. Died Oct. 25 of cardiac arrest.

Gilbert Beebower, 79. A co-author of a 1986 article demonstrating the superiority of asset allocation compared with market timing and stock picking, who worked at SEI Investments Co. in Oaks, Pennsylvania, from 1975 until his death. Died Oct. 25.

Lou Reed, 71. The New York-based rock musician, who co-founded the Velvet Underground and became one of rock music's most influential artists. Died Oct. 27 of complications from a liver transplant.

Leonard M. Leiman, 82. He led the securities-law practice at New York-based Reavis & McGrath when it merged in 1988 with Houston-based Fulbright & Jaworski, creating the seventh-largest U.S. law firm at the time. Died Oct. 30.

November

Walt Bellamy, 74. A member of the NBA Hall of Fame, he was one of only seven players to score more than 20,000 points and grab more than 14,000 rebounds. Died Nov. 2.

Rachel Benepe, 37. A U.S. protege of stock-picker Jean-Marie Eveillard at First Eagle Investment Management LLC, who managed its $1.5 billion First Eagle Gold Fund since 2009. Died Nov. 2 of cancer.

Charlie Trotter, 54. The Chicago-based chef who closed his namesake restaurant in 2012 after a 25-year run in which it won 11 James Beard Foundation Awards. Died Nov. 5.

Clarence “Ace” Parker, 101. Inducted into the Pro Football Hall of Fame in 1972, he played on New York teams in the 1940s, and twice spent the off-season playing baseball with the Philadelphia Athletics. Died Nov. 6.

Manfred Rommel, 84. He served as the former mayor of Stuttgart, Germany, for 22 years and was the son of Erwin Rommel, the German field marshal during World War II. Died Nov. 7 of Parkinson’s disease.

Sally Lloyd, 64. A third-generation banker who started her career in the early 1970s when few women worked on Wall Street and rose to managing director at Smith Barney. Died Nov. 11 of cancer.

John Tavener, 69. The U.K. composer best known for works such as “Song for Athene,” played at the funeral of Diana, Princess of Wales. Died Nov. 12.

Todd Christensen, 57. An NFL player from 1979 to 1988, who won two Super Bowl titles with the Oakland Raiders as a tight end and was voted All-Pro four times. Died Nov. 13 of complications from surgery.

Glafcos Clerides, 94. While president of Cyprus from 1993 to 2003, he oversaw the country’s entrance into the European Union in 2004. Died Nov. 15.

Doris Lessing, 94. The British author won the Nobel Prize in literature in 2007 and is best-known for “The Golden Notebook,” a story about an independent-minded woman growing up in Africa. Died Nov. 17.

G. Moffett Cochran, 63. The co-founder and CEO of New York-based Silvercrest Asset Management Group Inc., a firm serving wealthy families. Died Nov. 18 of cancer.

Michael Weiner, 51. As executive director of the Major League Baseball Players Association since 2009, he helped keep labor peace in the sport. Died Nov. 21 of cancer.

Peter B. Lewis, 80. The billionaire chairman of Progressive Corp., one of the biggest U.S. auto insurers, and a supporter of the medical use of marijuana. Died Nov. 23 of a heart attack at his home in Coconut Grove, Florida.

Robin Leigh-Pemberton, 86. He was Bank of England governor from 1983 to 1993. Died Nov. 24.

Matthew Bucksbaum, 87. The co-founder of General Growth Properties Inc., the second-biggest U.S. owner of shopping malls. Died Nov. 24 of respiratory failure.

Alfred Feld, 98. The longest-serving employee at Goldman Sachs, who joined the firm in 1933 and rose from office boy to private-wealth manager. Died Nov. 25.

Peter W. Kaplan, 59. The former editor of the New York Observer, which under his leadership chronicled the lives of New York's power elite and ran the column, "Sex and the City," which inspired a hit television series. Died Nov. 29 of cancer.

Paul Walker, 40. A Hollywood actor best-known for appearing in the "Fast and Furious" action movies. Died Nov. 30 of injuries as a passenger involved in a car crash.

December

Nelson Mandela, 95. The anti-apartheid freedom fighter, who endured 27 years in prison to become South Africa's first black president, then united the country and won the Nobel Peace Prize in 1993. Died Dec. 5 following a recurring lung infection.

Lawrence McCarthy, 49. Before becoming a senior managing director at Cantor Fitzgerald, he worked at Wasserstein Perella & Co., where he advised clients to sell Enron prior to its collapse, and at Lehman Brothers, where he warned colleagues in 2007 that the bank had taken on "far, far too much risk" by betting on the U.S. housing market. Died of an aneurysm on Dec. 11 in New York.

Peter O’Toole, 81. The British actor, who became an international star in 1962 for playing the lead in “Lawrence of Arabia” and received four Golden Globe awards and eight Oscar nominations. Died December 14.

Dennis Busti, 71. He was the CEO of corporate raider Saul Steinberg’s Reliance National Insurance Co., a unit created to handle high-risk insurance coverage for clients such as nuclear-plant operators. Died Dec. 14 at his home in Eastchester, New York.

Joan Fontaine, 96. Born in Tokyo to British parents, the actress spent most of her life in the U.S. and won an Academy Award for best actress for her performance in the 1941 Alfred Hitchcock film “Suspicion,” beating her sister, Olivia de Havilland, for the honor. Died Dec. 15.

Graham Mackay, 64. The former CEO of London-based SABMiller Plc,
who built the company into the world’s second-biggest brewer and acquired Australia’s Foster’s Group Ltd. in 2011 and Miller Brewing Co., a U.S. beer maker, in 2002. Died Dec. 18.

Ronnie Biggs, 84. He helped stage Britain’s Great Train Robbery in 1963, escaped from prison and eluded Scotland Yard for 36 years before giving himself up in 2001. Died Dec. 18 after a series of strokes.

Al Goldstein, 77. A Brooklyn-born pornographer who published Screw magazine, hosted a public access cable-TV show in New York during the city’s sleazy days in the 1970s, before Times Square was cleaned up and drawing families to “The Lion King.” Died Dec. 19.

Sergio Loro Piana, 65. The Italian cashmere clothier, who along with his brother, Pier Luigi Loro Piana, became billionaires after selling 80 percent of their company, Loro Piana SpA, to Paris-based LVMH Moet Hennessy Louis Vuitton SA. Died Dec. 19.

John S.D. Eisenhower, 91. The son of former U.S. President Dwight Eisenhower, he was a brigadier general in the U.S. Army Reserve, wrote books on military history and was appointed ambassador to Belgium by President Richard Nixon in 1969. Died Dec. 21.

Edgar M. Bronfman, 84. The Canadian-born second-generation heir who expanded the Seagram Co. with oil, gas and chemical investments and served as president of the World Jewish Congress from 1981 to 2007. Died Dec. 21 at his home in New York.

Mikhail Kalashnikov, 94. He was the Russian inventor of what would become the world's most popular assault rifle, the AK-47. Died Dec. 23.

Robert W. Wilson, 87. He founded a New York-based hedge fund, amassed a net worth of about $800 million and gave most of it to charities, primarily conservation groups. Died Dec. 23 of suicide.

--Editors: Charles W. Stevens, David Henry

http://www.bloomberg.com/news/2013-12-26/thatcher-mandela-chavez-are-among-notable-deaths-in-2013.html

IN THE CIRCUIT COURT OF THE FIFTEENTH
JUDICIAL CIRCUIT, IN AND FOR PALM
BEACH COUNTY, FLORIDA

CASE NO.: 502009CA040800XXXXMBAG

JEFFREY EPSTEIN,

Plaintiff,

vs.

SCOTT ROTHSTEIN, individually, BRADLEY J.
EDWARDS, individually, and L.M., individually,

Defendant,
_____________________________________/

DEFENDANT/COUNTER-PLAINTIFF'S RESPONSE IN OPPOSITION TO PLAINTIFF/COUNTER-DEFENDANT'S MOTION FOR SUMMARY JUDGMENT ON DEFENDANT/COUNTER-PLAINTIFF'S FOURTH AMENDED COUNTERCLAIM

Defendant/Counter-Plaintiff Bradley J. Edwards, by and through his undersigned counsel, hereby submits this Response in Opposition to Plaintiff/Counter-Defendant Jeffrey Epstein's Motion for Summary Judgment. Epstein seeks Summary Judgment on the claims of abuse of process and malicious prosecution set forth in Brad Edwards' Fourth Amended Counterclaim. Each of the grounds asserted in support of Epstein's Motion for Summary Judgment are without merit and must be denied.

In Epstein's Amended Complaint he carries forth the essence of all claims asserted in his original Complaint. In that pleading Epstein essentially alleges that Edwards joined Rothstein in the abusive prosecution of sexual assault cases against Epstein to "pump" the cases to Ponzi scheme investors. The purported "proof" of the allegations against Edwards, as referenced in the Second Amended Complaint and in Epstein's Motion for Summary Judgment, includes Edwards' alleged contacts with the media, his attempts to obtain discovery from high profile persons with whom Epstein socialized, press reports of Rothstein's known illegal activities, the use of "ridiculously inflammatory" language and arguments in court. But as the evidence submitted in opposition to Epstein's Motion for Summary Judgment reflects, Epstein filed his claims and continued to pursue claims despite his knowledge that his claims could never be successful because they were both false and unsupported by any reasonable belief of suspicion that they were true. Epstein knew that he had in fact molested each of the minors represented by Brad Edwards. He also knew that each litigation decision by Brad Edwards was grounded in proper litigation judgment about the need to pursue effective discovery against Epstein, particularly in the face of Epstein's stonewalling tactics. Epstein also knew that he suffered no legally cognizable injury proximately caused by the falsely alleged wrongdoing on the part of Edwards. Moreover, Epstein had no intention of waiving his Fifth Amendment privilege against self-incrimination in order to avoid providing relevant and material discovery that Epstein would need in the course of prosecuting his claims and to which Edwards was entitled in defending those claims. Epstein knew that his prosecution of his claims would be barred by the sword-shield doctrine. Most significantly, the evidence submitted in the supporting papers would compel a fact finder to determine that Epstein had no basis in law or in fact to pursue his claims against Edwards and that Epstein was motivated by a single ulterior motive to attempt to intimidate Edwards and his clients and others into abandoning or settling their legitimate claims for less than their just and reasonable value. The evidence demonstrates that Epstein did not file these claims for the purpose of collecting money damages since he knew that he never suffered any damage as a consequence of any alleged wrongdoing by Edwards but filed the claim to require Edwards to expend time, energy and resources on his own defense, to embarrass Edwards and impugn his integrity and deter others with legitimate claims against Epstein from pursuing those claims. Indeed, the evidence demonstrates that Epstein continued to pursue his claims by filing the Second Amended Complaint alleging abuse of process against Edwards *even after* he had paid significant sums in settlement of the claims instituted by Mr. Edwards' clients against Mr. Epstein.[1]

---

[1] The evidence marshalled in support of these assertions is set forth in the previously filed documents in this Court. Those documents include Exhibit "A" – Edwards' Statement of Undisputed Facts; Exhibit "B" – Edwards' Renewed Motion for Summary Judgment; Exhibit "C" – Edwards' October 19, 2012 Second Renewed Motion for Leave to Assert Claim for Punitive Damages; Exhibit "D" – Edwards' Notice of Filing of Transcript of Telephone Interview of Virginia Roberts in Support of Motion for Leave to Amended to Assert Punitive Damages; Exhibit "E" – Transcript of Deposition of Jeffrey Epstein dated January 25, 2012; Exhibit "F' – Deposition of Bradley Edwards dated March 23, 2010; Exhibit "G" - Deposition of Scott Rothstein dated June 14, 2012; Exhibit "H" – Order of

The record reflects that on the eve of the hearing of Edwards' Motion for Summary Judgment directed to the Second Amended Complaint and in light of the compelling evidence of the lack of any wrongdoing on the part of Mr. Edwards, the sole remaining abuse of process claim was dismissed by Epstein.

As discussed, *infra* each of the grounds asserted by Epstein in this Motion for Summary Judgment must be rejected. The litigation privilege does not serve as a bar to the prosecution of Edwards' claims against Epstein. Moreover, the evidence submitted by Edwards supports each of the elements of the claims asserted by Edwards against Epstein which are identified in Epstein's Motion.

Response to Epstein's Statement of Undisputed Facts

The evidence marshalled by Edwards in support of his claims against Epstein which are referenced in footnote 1 mandates the conclusion that, at a minimum, disputed facts exist with respect to the elements of each claim addressed by Epstein in his Motion. The facts presented in the various papers would allow the jury to make a determination that Epstein knew that Brad Edwards properly exercised his legitimate judgment regarding the need to pursue proper and effective discovery against him to support the claims which Epstein knew were legitimate. That evidence, referenced herein, further demonstrated that Epstein filed his claims without probable cause and further that there was a bonafide termination in favor of Edwards. That evidence further demonstrates that the elements of the claim of abusive process have been established.

The following additional comments are directed at some of the key purported "undisputed" material facts asserted by Epstein, especially those referenced in his Memorandum of Law. Also set forth are key evidentiary matters which undermine Epstein's contentions and which support the proposition that material issues of fact exist which compel the denial of the Motion for Summary Judgment.

Judge Crow dated March 29, 2012; Exhibit "I" – Deposition of Bradley Edwards dated October 10, 2013; Exhibit "J" – Deposition of Bradley Edwards dated May 15, 2013.

None of the public materials identified by Epstein in his Motion make reference to any wrongdoing by Brad Edwards. Rather, Epstein seeks to pyramid one impermissible inference upon another from his citation to these materials to support his otherwise unsubstantiated and non-verifiable conclusion that he had sufficient evidence to proceed with claims of wrongdoing against Edwards. In truth, as reflected in Edwards' deposition and his supplemental affidavit, he has no involvement in any fraud perpetrated by Rothstein (Edwards' deposition of March 23, 2010 at 301-302; Edwards Affidavit attached to Statement of Undisputed Facts as Exhibit "N" at paragraphs 8-10, paragraph 20, paragraphs 22-23; Exhibit "H" – Deposition of Scott Rothstein at pp. 62-63, 114, and 121-124). Therefore, any allegations relating to Rothstein's activities simply have no bearing on the legitimacy of any of the claims against Edwards. Edwards could not have possibly "pumped" cases to investors when he never participated in any communications with investors. Rather, Edwards had a duty to his clients to zealously pursue discovery to achieve a maximum recovery against Epstein. Edwards cannot be liable for taking appropriate action that his ethical duties as an attorney required. The evidence also reflects that Edwards filed all three of his cases almost a year before he was hired by RRA or even knew Scott Rothstein (Edwards' Affidavit, Exhibit "N" attached to Statement of Undisputed Facts). The language set forth in his Complaints remain virtually unchanged from the first filing in 2008 and, as the evidence shows, the claims asserted against Epstein from the outset were true. The citation to public documents is a convenient ruse; Epstein was not only liable for the molestation of the clients of Brad Edwards, he was also a serial molester of minors – even as young as twelve years of age (Exhibit "A" – Edwards' Statement of Undisputed Material Facts paragraphs 1-43; Exhibit "D" – Statement of Virginia Roberts pp. 16-17). Epstein entered a plea of guilty to felony charges involving prostitution and the solicitation of a minor for the purposes of prostitution (Exhibit "E" – Deposition of Jeffrey Epstein, March 17, 2010, pp. 101-103). Epstein also entered into an agreement with the United States Attorney's Office acknowledging that approximately 34 other young girls could receive payments from him under the

Federal Statute providing for compensation to victims of child abuse.. (Exhibit "A" – Edwards' Statement of Undisputed Material Facts, paragraphs 41-43).

On July 6, 2010 Epstein ultimately paid to settle all three of the cases Edwards had filed against him (Exhibit "A" – Edwards' Statement of Undisputed Material Facts, paragraphs 84-85). At Epstein's request, the terms of the settlement were kept confidential. The sum that he paid to settle all these cases in therefore not filed with this pleading and will be provided to the court for in camera review. Epstein chose to make this payment as a result of a Federal Court ordered mediation process which he himself sought. Epstein entered into the settlements in July 2010 more than seven months after he filed his lawsuit against Edwards and before he filed his Second Amended Complaint alleging abuse of process on August 22, 2011.

Further, Epstein could not have been the victim of any scheme to pump the cases against him because he never paid to settle the cases until well after Edwards had left RRA and severed all connection with Rothstein in December 2009 (Edwards' Affidavit attached to Statement of Undisputed Facts as Exhibit "N," paragraph 20). Moreover, Epstein could not have suffered any damage as a result of the perpetration of the Ponzi scheme by Rothstein because he was not an investor in the scheme.

Perhaps the most significant evidence presented in opposition to Epstein's Motion for Summary Judgment is the telephone interview of Virginia Roberts submitted in Support of Edwards' Motion for Punitive Damages (Exhibit "D"). In addition to the specious claims against Edwards relating to his alleged involvement in a Ponzi scheme, Epstein, in asserting his claims, primarily relied upon the pursuit by Edwards of testimony from his close friends and associates (See Second Amended Complaint, paragraph 32, pp. 11-13). Reliance on these assertions is also threaded through Epstein's Motion for Summary Judgment in his citation to the public documents referencing the pursuit of such discovery. But as set forth in detail in Edwards' Motion for Final Summary Judgment (Exhibit "B") at pages 14-16, that discovery was entirely appropriate and Epstein knew it. Specifically, as reflected in the statement of undisputed facts submitted by Mr. Edwards in support of his Motion for Summary Judgment, Edwards had a sound legal basis for believing that Donald Trump, Allen Dershowitz, Bill Clinton, Tommy Mattola, David Copperfield and Governor Bill Richardson had relevant and discoverable information (Exhibit "A" – Edwards' Statement of Undisputed Facts, paragraphs 69-81). That belief was reinforced by the testimony of Virginia Roberts (Exhibit "D" pp. 10-17, 21-23). Epstein's assertion of impropriety in the pursuit of this discovery clearly evidences his bad faith attempts to attribute wrongdoing to Edwards when he knew, in fact, that the pursuit of that discovery was entirely appropriate under the circumstances of this case.

Finally, any attempt by Epstein to rely upon what he claims are undisputed facts to support his Motion for Summary Judgment are undermined by his refusal to provide any testimony on the key issues and evidence which would demonstrate the validity and strength of each of the claims brought against him by Brad Edwards. Epstein's depositions of March 17, 2010 and January 25, 2012 were replete with refusals of Epstein to testify based upon his Fifth Amendment privilege. Questions that Epstein refused to answer in his depositions and the reasonable inferences that a fact finder would draw and which would otherwise bear on the arguments submitted by Epstein in support of his Motion for Summary Judgment are as follows:

- Question not answered: "I want to know whether you have any knowledge of evidence that Bradley Edwards personally ever participated in devising a plan through which were sold purported confidential assignments of a structured payout settlement?" Reasonable inference: No knowledge that Brad Edwards ever participated in the Ponzi scheme.

- Question not answered: "Specifically what are the allegations against you which you contend Mr. Edwards ginned up?" Reasonable inference: No allegations against Epstein were ginned up.

- Question not answered: "Well, which of Mr. Edwards' cases do you contend were fabricated?" Reasonable inference: No cases filed by Edwards against Epstein were fabricated.

- Question not answered: "Did sexual assaults ever take place on a private airplane on which you were a passenger?" Reasonable inference: Epstein was on a private airplane while sexual assaults were taking place.

- Question not answered: "How many minors have you procured for prostitution?" Reasonable inference: Epstein has procured multiple minors for prostitution.

- Question not answered: "Is there anything in L.M.'s Complaint that was filed against you in September of 2008 which you contend to be false?" Reasonable inference: Nothing in L.M.'s complaint filed in September of 2008 was false – i.e., as alleged in L.M.'s complaint, Epstein repeatedly sexually assaulted her while she was a minor and she was entitled to substantial compensatory and punitive damages as a result.

- Question not answered: "I would like to know whether you ever had any physical contact with the person referred to as Jane Doe in that [federal] complaint?" Reasonable inference: Epstein had physical contact with minor Jane Doe as alleged in her federal complaint.

- Question not answered: "Did you ever have any physical contact with E.W.?" Reasonable inference: Epstein had physical contact with minor E.W. as alleged in her complaint.

- Question not answered: "What is the actual value that you contend the claim of E.W. against you has?" Reasonable inference: E.W.'s claim against Epstein had substantial actual value.

  (See Exhibit "A" – Edwards' Statement of Undisputed Material Facts, paragraphs 93-120 for page references.)

A jury could conclude, therefore, from the adverse inferences drawn against Epstein that he was liable for the claims brought by Brad Edwards and that he had no basis for the pursuit of his efforts to intimidate and extort Edwards and his clients in the pursuit of those claims.

The Litigation Privilege Does Not Bar the Claims of Abuse of Process and Malicious Prosecution

Epstein contends he is entitled to absolute immunity pursuant to the litigation privilege as to both claims asserted by Edwards because all actions taken by him occurred during the litigation of his abuse of process claim against Edwards. For support, he relies primarily on the decision of *Wolfe v. Foreman*, 2013 WL 3724763 (Fla. 3d DCA July 17, 2013), wherein the Third District found that the litigation privilege barred both an abuse of process claim and a malicious prosecution cause of action. *Wolfe* is still on rehearing and, thus, is not a final opinion. As a result, it is not binding, nor persuasive. Moreover, *Wolfe* undercuts the long-standing recognition of the viability of a claim for malicious prosecution in its own District and other Florida state and federal courts. *See, SCI Funeral Svs. of Fla., Inc. v. Henry*, 839 So. 2d 702, n.4 (Fla. 3d DCA 2002) ("As the *Levin* court cited *Wright v. Yurko*, 446 So. 2d 1162, 1165 (Fla. 5th DCA, 1984), with approval, presumably the cause of action for malicious prosecution continues to exist and would not be barred by the litigation privilege."); *Boca Investors Group, Inc. v. Potash*, 835 So. 2d 273, 275 (Fla. 3d DCA 2002) (Cope, J., concurring) (litigation privilege would not be a bar to a malicious prosecution action); *North Star Capital Acquisitions, LLC v. Krig*, 611 Fed. Supp. 2d 1324 (M.D. Fla. 2009) ("However, not every event bearing any relation to litigation is protected by the privileged because,... "if the litigation privilege applied to all actions preliminary to or during judicial proceedings, an abuse of process claim would never exist, nor would a claim for malicious prosecution"); *Cruz v. Angelides*, 574 So. 2d 278 (Fla. 3d DCA 1991)("the law is well settled that a witness in a judicial proceeding,... is absolutely immune from any civil liability, save perhaps malicious prosecution, for testimony or other sworn statements which he or she gives in the course of the subject proceeding."); *Johnson v. Libow*, 2012 WL 4068409 (Fla. 15th Jud. Cir. March 1, 2012)(the purpose of the litigation privilege does not preclude the tort of malicious prosecution).

In *Wright v. Yurko supra,* the Fifth District Court of Appeal rejected the application of the litigation privilege to a malicious prosecution action brought by a physician against his patients and an expert after he successfully defended a malpractice claim. Also of significance is the Second District's opinion in *Olson v. Johnson,* 961 So. 2d 351 (Fla. 2d DCA 2007). In that case, the court observed that the litigation (or judicial) privilege would not apply to bar a malicious prosecution action which arose as a result of a false accusation of criminal liability where the prosecution was based, in part, on the testimony of the defendants in the criminal case. The court ruled that the privilege (either absolute or qualified) which might otherwise apply to a defamation claim for statements made during the course of a judicial proceeding did not bar a malicious prosecution claim.

In light of the implicit recognition by the Supreme Court in *Levin* that a claim of malicious prosecution is not barred by the litigation privilege – an implicit recognition acknowledged by the Third District itself – Epstein's reliance on *Wolfe* is misplaced. *Wolfe* is also factually distinguishable from Edwards' claims against Epstein. *Wolfe* involved a malicious prosecution action against attorneys. Separate policy considerations might serve to impose additional limitations on the assertion of malicious prosecution claims against attorneys – against whom alternative remedies exist such as bar disciplinary proceedings. *See Taylor v. McNichols*, 243 P.2d 642 (Idaho 2010). Moreover, in light of the decisions in *Wright v. Yurko, supra* and *Olson v. Johnson*, *supra*, the weight of authority supports the proposition that the litigation privilege would not apply to malicious prosecution claims.

Both the Third and Fourth Districts have applied the litigation privilege to abuse of process claims. However, *Wolfe* itself, and the decisions of the Third and Fourth Districts cited in Wolfe, involved the litigation privilege as applied to claims of abuse of process by attorneys. None of the cases involved the extraordinary actions of an individual party like Epstein who carried out a course of action against Plaintiff's counsel with a singular purpose unrelated to any legitimate judicial goal. Under the compelling facts of this case, where the actions of Epstein are coupled with the elements of malice and absence of probable cause arising from the unfounded filing of the claims against Edwards, the litigation privilege should not have any applicability to the abuse of process claim asserted by Edwards.

There are Disputed Issues of Fact Precluding Summary Judgment on the Abuse of Process Claim

An abuse of process claim requires pleading and proof of the following three elements: 1) that the defendant made an illegal, improper or perverted use of process; 2) that the defendant had ulterior motives or purposes in exercising such illegal, improper, or perverted use of process; and 3) that, as a result of such action on the part of the defendant, the plaintiff suffered damage." *See S&I Invs. v. Payless*

Flea Mkt., 36 So. 3d 909, 917 (Fla. 4th DCA 2010)(citation omitted). The case law is clear that on an abuse of process claim a "plaintiff must prove that the process was used for an immediate purpose other than that for which it was designed." *Id.* (citation omitted). Where the actions taken by a party in a particular lawsuit are designed to coerce another into taking some collateral action not properly involved in the proceeding a claim of abuse of process is stated. *Miami Herald Publishing Company v. Ferre,* 8636 F. Supp. 970 (S.D. Fla. 1985).

In a case for abuse of process, the question of whether the plaintiff's case satisfies the requisite elements is largely a question for a jury. *See* Patrick John McGinley, 21 Fla. Prac., Elements of an Action § 50:1 (2013-2014 ed.)(citing *Gatto v. Publix Supermarket, Inc.*, 387 So. 2d 377 (Fla. 3d DCA 1980)).

The usual case of abuse of process involves some form of extortion. *Scozari v. Barone*, 546 So. 2d 750, 751(Fla. 3d DCA 1989) (citing *Bothmann v. Harrington,* 458 So. 2d 1163, 1169 (Fla. 3d DCA 1984)). That is *exactly* what has transpired here. Epstein employed the extraordinary financial resources at his disposal to intimidate his molestation victims and Edwards into abandoning their legitimate claims or resolving those claims for substantially less than their just and reasonable value. Consequently, since Epstein's sole purpose and ulterior motive for filing the complaint without probable cause was in an effort to extort, to wit: to force his molestation victims and Edwards to settle for minimal amounts, that filing and everything subsequently done to pursue the claims constitutes an abuse of process. *See* Exhs. A at 18-27, C at 4-7. Because Edwards has conclusively demonstrated that Epstein's actions in pursuing his claims were designed to coerce Edwards (and his client) to take some collateral action not properly involved in the proceedings and did so with an ulterior purpose, summary judgment directed at the abuse of process claim must fail. The damages suffered by Edwards include: (a) injury to his reputation; (b) mental anguish, embarrassment and anxiety; (c) fear physical injury to himself and members of his family; (d) the loss of the value of his time required to be diverted from his professional responsibility; and (e) the cost of defending against Epstein's spurious and baseless claims. All the elements of the claim for abuse of process have been satisfied. This case, then, falls within the parameters of the Third District's Decision in *Scozari v. Barone, supra* in which the court reversed the entry of summary judgment for the defendant on claims of malicious prosecution and abuse of process. With respect to the abuse of process claim, the court stated that "if there was no reasonable basis in law and fact to bring the action to impress a lien on property, and this was done without any reasonable justification under law and to force or compel the appellant to resolve some custody dispute, induce the appellant to pay money, or tie up the appellant's property, then there has been an abuse of process." *Id* at 752.

There are Disputed Issues of Fact Precluding Summary Judgment on the Claim of Malicious Prosecution

Here, Epstein's voluntary dismissal of his abuse of process claims against Edwards amounted to a bona fide termination of the proceedings. He knew his allegations were unsupported by evidence (See discussion above at pages 3-6). Knowing he lacked *any* verifiable evidence against Edwards, on the eve of the summary judgment hearing, Epstein effectively conceded that fact by voluntarily dismissing his claims. Hence, it is evident that Epstein took voluntary dismissal of his claims because he knew he did not have probable cause or an evidentiary basis to support the allegations. *See Cohen v. Corwin*, 980 So. 2d 1153 at 1156 (citing *Union Oil of California, Amsco Division v. Watson,* 468 So. 2d 349 at 354 (stating that "where a dismissal is taken because of insufficiency of the evidence, the requirement of a favorable termination is met")). Accordingly, the manner of termination reflects on the merits of the case and there was a bona fide termination of Epstein's civil proceeding against Edwards (See Judge Crow's Order of March 29, 2012 denying Motion to Dismiss re: Issue of Bonafide Termination attached as Exhibit "H").

Epstein's only other issue with Edwards' counterclaim for malicious prosecution is that he did not lack probable cause in pursing his claims against Edwards. As established by the record, Epstein did, in fact, lack probable cause to assert his claims against Edwards (See discussion above). Epstein's purported reliance on public filings, including the Scherer Complaint against Rothstein is unavailing. As discussed above, the evidence warrants the finding that Epstein knew that Edwards was legitimately pursuing the claims on behalf of his clients which included the effort to secure testimony from Epstein's close confidants. Therefore, Epstein cannot rely upon the referenced public documents to support his claims against Edwards given that he knows that information to be untrue and he refuses to answer questions about the veracity of the information. *See* Exh. G at pgs. 53:6-24; 78:16-24; 87:20-88:14. Consequently, Epstein had no good faith basis to rely on such information.

Epstein's Assertion of his Fifth Amendment Privilege Gives Rise to Adverse Inferences Pertinent to His Motion for Summary Judgment and Precludes His Reliance on Purported Undisputed Facts

As discussed above, Epstein's multiple invocations of his Fifth Amendment Privilege results in adverse inferences which directly impact the issues advanced in his Motion for Summary Judgment. "It is well settled that the Fifth Amendment does not forbid adverse inferences against parties to civil actions when they refuse to testify in response to probative evidence offered against them." *Baxter v. Palmigiano,* 425 U.S. 308, 318 (1976); *Accord, Vasquez v. State,* 777 So. 2d 1200, 1203 (Fla. at 2001). The reason for this rule "is both logical and utilitarian. A party may not trample upon the rights of others and then escape the consequences by invoking a constitutional privilege – at least not in a civil setting." *Fraser v. Security and INV. Corp,* 615 So. 2d. 841, 842 (Fla. 4th DCA 1993). The adverse inferences drawn from Epstein's assertion of the Fifth Amendment undercut his claim of justifiable reliance based upon the purported undisputed material facts to support his Motion for Summary Judgment.

Moreover, because Epstein elected to hide behind the shield of his right against self-incrimination to preclude his disclosing any relevant information about the criminal activity at the center of his claims, he was effectively barred from prosecuting his abuse of process claim against Edwards. Similarly, Epstein should be barred from utilizing the Fifth Amendment privilege to secure summary judgment based upon assertions of fundamental facts when Epstein refused to testify on essential issues pertinent to the arguments advanced in support of his Motion for Summary Judgment. Under the well-established "sword and shield" doctrine, Epstein could not seek damages from Edwards while at the same time asserting a Fifth Amendment privilege to block relevant discovery. *See* Exhs. B at 14-21, C at 18-25, G at 53:6-24; 78:16-24; 87:20-88:14. The same policies which underlie the sword and shield doctrine as applied to the recovery of affirmative relief should also apply to attempts to advance positions with respect to a Motion for Summary Judgment which would have the effect of securing relief against certain claims.

"[T]he law is well settled that a plaintiff is not entitled to both his silence and his lawsuit." *Boys & Girls Clubs of Marion County, Inc. v. J.A.*, 22 So. 3d 855, 856 (Fla. 5th DCA 2009)(Griffin, J., concurring specially). Thus, "a person may not seek affirmative relief in a civil action and then invoke the fifth amendment to avoid giving discovery, using the fifth amendment as both a 'sword and a shield.'" *DePalma v. DePalma,* 538 So. 2d 1290, 1290 (Fla. 4th DCA 1989)(quoting *DeLisi v. Bankers Insurance Co.,* 436 So. 2d 1099 (Fla. 4th DCA 1983)). Put another way, "[a] civil litigant's fifth amendment right to avoid self-incrimination may be used as a shield but not a sword. This means that a plaintiff seeking affirmative relief in a civil action may not invoke the fifth amendment and refuse to comply with the defendant's discovery requests, thereby thwarting the defendant's defenses." *Rollins Burdick Hunter of New York, Inc. v. Euroclassic Limited, Inc.*, 502 So. 2d 959 (Fla. 3d DCA 1983).. For the same reasons, Epstein should be precluded from advancing arguments based on purported statements of undisputed fact which cannot be effectively challenged in light of his assertion of the Fifth Amendment. Epstein has done precisely what well-established law prohibits.

Conclusion

Based upon the foregoing, the Defendant, Counter-Plaintiff, Bradley Edwards respectfully submits that Jeffrey Epstein's Motion for Summary Judgment must be denied.

I HEREBY CERTIFY that a true and correct copy of the foregoing was sent via E-Serve to all Counsel on the attached list, this 17th day of January, 2014.

WILLIAM B. KING
Florida Bar No.: 181773
Attorney E-Mail: wbk@searcylaw.com and
kar@searcylaw.com
Primary E-Mail: eservice@searcylaw.com
Secondary E-Mail: _ScarolaTeam@searcylaw.com
Searcy Denney Scarola Barnhart & Shipley, P.A.
2139 Palm Beach Lakes Boulevard
West Palm Beach, Florida 33409
Phone: (561) 686-6300
Fax: (561) 383-9456
Attorney for Bradley J. Edwards

COUNSEL LIST

William Chester Brewer, Esquire
wcblaw@aol.com; wcbcg@aol.com
250 S Australian Avenue, Suite 1400
West Palm Beach, FL 33401
Phone: (561)-655-4777
Fax: (561)-835-8691
Attorneys for Jeffrey Epstein

Jack A. Goldberger, Esquire
jgoldberger@agwpa.com;
smahoney@agwpa.com
Atterbury, Goldberger & Weiss, P.A.
250 Australian Avenue South, Suite 1400
West Palm Beach, FL 33401
Phone: (561)-659-8300
Fax: (561)-835-8691
Attorneys for Jeffrey Epstein

Bradley J. Edwards, Esquire
staff.efile@pathtojustice.com
Farmer, Jaffe, Weissing, Edwards, Fistos &
Lehrman, FL
425 North Andrews Avenue, Suite 2
Fort Lauderdale, FL 33301
Phone: (954)-524-2820
Fax: (954)-524-2822
Attorneys for Jeffrey Epstein

Fred Haddad, Esquire
Dee@FredHaddadLaw.com;
haddadfm@aol.com; fred@fredhaddadlaw.com
Fred Haddad, P.A.
One Financial Plaza, Suite 2612
Fort Lauderdale, FL 33394
Phone: (954)-467-6767
Fax: (954)-467-3599
Attorneys for Jeffrey Epstein

Marc S. Nurik, Esquire
marc@nuriklaw.com
Law Offices of Marc S. Nurik
One E Broward Blvd., Suite 700
Fort Lauderdale, FL 33301
Phone: (954)-745-5849
Fax: (954)-745-3556
Attorneys for Scott Rothstein

Tonja Haddad Coleman, Esquire
tonja@tonjahaddad.com;
Debbie@Tonjahaddad.com;
efiling@tonjahaddad.com
Tonja Haddad, P.A.
315 SE 7th Street, Suite 301
Fort Lauderdale, FL 33301
Phone: (954)-467-1223
Fax: (954)-337-3716
Attorneys for Jeffrey Epstein

IN THE CIRCUIT COURT OF THE 15TH
JUDICIAL CIRCUIT IN AND FOR PALM
BEACH COUNTY, FLORIDA

Case No.:50 2009 CA 040800XXXXMBAG

JEFFREY EPSTEIN,

Plaintiff,

vs.

SCOTT ROTHSTEIN, individually, and
BRADLEY J. EDWARDS, individually,

Defendants,
_________________________________/

STATEMENT OF UNDISPUTED FACTS

Defendant Bradley J. Edwards, Esq., offers the following specific facts as the undisputed material facts in this case. Each of the following facts is numbered separately and individually to facilitate Epstein's required compliance with Fla. R. Civ. P. 1.510(c) ("The adverse party shall identify . . . any summary judgment evidence on which the adverse party relies."). All referenced exhibits and attachments have previously been filed with the Court and provided to Epstein.

Sexual Abuse of Children By Epstein

1. Defendant Epstein has a sexual preference for young children. Deposition of Jeffrey Epstein, Mar. 17, 2010, at 110 (hereinafter "Epstein Depo.") (Deposition Attachment #1).[1]

---

[1] When questioned about this subject at his deposition, Epstein invoked his Fifth Amendment right to remain silent rather than make an incriminating admission. Accordingly, Edwards is entitled to the adverse inference against Epstein that, had Epstein answered, the answer would have been unfavorable to him. "[I]t is well-settled that the Fifth Amendment does not forbid adverse inferences against parties to

EXHIBIT

A

2. Epstein repeatedly sexually assaulted more than forty (40) young girls on numerous occasions between 2002 and 2005 in his mansion in West Palm Beach, Florida. These sexual assaults included vaginal penetration. Epstein abused many of the girls dozens if not hundreds of times. Epstein Depo. at 109 ("Q: How many times have you engaged in oral sex with females under the age of 18?" A: [Invocation of the Fifth Amendment]); Deposition of Jane Doe, September 24, 2009 and continued March 11, 2010, at 527 (minor girl sexually abused at least 17 times by Epstein) (hereinafter "Jane Doe Depo") (Deposition Attachment #2); *id.* 564-67 (vaginal penetration by Epstein with his finger), 568 (vaginal penetration by Epstein with a massager); Deposition of L.M., September 24, 2009, at 73 (hereinafter "L.M. Depo") (Deposition Attachment #3) (describing the manner in which Epstein abused her beginning when LM was 13 years old, touching her vagina with his fingers and vibrator) at 74, line 12-13 (she was personally molested by Epstein more than 50 times), at 164, line 19-23 and 141, line 12-13 and 605, line 3-6 (describing that in addition to being personally molested by Epstein she was paid $200 per underage girl she brought Epstein and she brought him more than seventy (70) underage girls - she told him that she did not want to bring him any more girls and he insisted that she continue to bring him underage girls); Deposition of E.W., May 6, 2010 (hereinafter "E.W. Depo") (Deposition Attachment #4) at 115-116, 131 and 255 (describing Epstein's abuse of her beginning at age 14 when he paid her for touching her vagina, inserting his fingers and

---

civil actions when they refuse to testify in response to probative evidence offered against them." *Baxter v. Palmigiano,* 425 U.S. 308, 318 (1976); *accord Vasquez v. State*, 777 So.2d 1200, 1203 (Fla. App. 2001). The reason for this rule "is both logical and utilitarian. A party may not trample upon the rights of others and then escape the consequences by invoking a constitutional privilege – at least not in a civil setting." *Fraser v. Security and Inv. Corp.*, 615 So.2d 841, 842 (Fla. App. 1993).

using a vibrator and he also paid her $200 for each other underage female E.W. brought him to molest. She brought him between 20 and 30 underage females); Deposition of Jane Doe #4, date (hereinafter "Jane Doe #4 Depo") (Deposition Attachment #5) at 32-34, and 136 (she describes first being taken to Epstein at 15 years old, "Being fingered by him, having him use a vibrator on [me], grabbing my nipples, smelling my butt, jerking off in front of me, licking my clit, several times.").

3. At all relevant times Edwards has had a good faith basis to conclude and did conclude[2] that Epstein was able to access a large number of underage girls through a pyramid abuse scheme in which he paid underage victims $200-$300 cash for each other underage victim that she brought to him. *See* Palm Beach Police Incident Report at 87 (hereinafter "Incident Report") (Exhibit "A").[3] The Palm Beach Police Incident Report details Epstein's scheme for molesting underage females. Among other things, the Incident Report outlines some of the experiences of other Epstein victims. When S.G, a 14 year old minor at the time, was brought to Epstein's home, she was taken upstairs by a woman she believed to be Epstein's assistant. The woman started to fix up the room, putting covers on the massage table and bringing lotions out. The "assistant" then left the room and told S.G. that Epstein would be up in a second. Epstein walked over to S.G. and told her to take her clothes off in a stern voice. S.G. states in the report she did not know what to do, as she was the only one there. S.G. took off her shirt, leaving her bra on. Epstein, then in a towel told her to take off everything. S.G. removed her pants leaving

---

[2] In support of all assertions concerning the actions Edwards took, what Edwards learned in the course of his representation of his clients, Edwards's good faith beliefs and the foundation for those beliefs, see Edwards Affidavit and specifically paragraphs 25 and 25 of that Affidavit.

[3] For clarity, depositions attached to this memorandum will be identified numerically as attachments #1, #2, #3, etc., while exhibits attached to this memorandum will be identified alphabetically as exhibits A, B, C, etc.

on her thong panties. Epstein then instructed S.G to give him a massage. As S.G gave Epstein a massage, Epstein turned around and masturbated. S.G. was so disgusted, she did not say anything; Epstein told her she "had a really hot body." *Id.* at 14. In the report, S.G. admitted seeing Jeffrey Epstein's penis and stated she thought Epstein was on steroids because he was a "really built guy and his wee wee was very tiny." *Id. at* 15.

4. The exact number of minor girls who Epstein assaulted is known only to Epstein. However, Edwards had a good faith basis to believe and did in fact believe that Epstein's victims were substantially more than forty (40) in number. In addition to the deposition excerpts from two of his many victims above about the number of underage girls brought to Epstein and the Palm Beach incident report, there is overwhelming proof that the number of underage girls molested by Epstein through his scheme was in the hundreds. *See* Complaint, Jane Doe 102 v. Epstein, (hereinafter Jane Doe 102 complaint) (Exhibit "B"); *see also* Deposition of Jeffrey Epstein, April 14, 2010, at 442, 443, and 444 (Epstein invoking the 5th on questions about his daily abuse and molestation of children) (Deposition Attachment #6).

5. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that Epstein and his attorneys knew of the seriousness of the criminal investigation against him and corresponded constantly with the United States Attorney's Office in an attempt to avoid the filing of numerous federal felony offenses, which effort was successful. *See* Correspondence from U.S. Attorney's Office to Epstein (hereinafter "U.S. Attorney's Correspondence") (Composite Exhibit "C) (provided in discovery during the Jane Doe v. Epstein case).

6. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that, more specifically, Epstein's attorneys knew of Epstein's scheme to recruit minors for sex and also knew that these minors had civil actions that they could bring against him. In fact, there was much communication between Epstein's attorneys and the United States Prosecutors in a joint attempt to minimize Epstein's civil exposure. For example, on October 3, 2007, Assistant U.S. Attorney Marie Villafaña sent an email (attached hereto as Exhibit "D") to Jay Lefkowitz, counsel for Epstein, with attached proposed letter to special master regarding handling numerous expected civil claims against Epstein. The letter reads in pertinent part,

> "The undersigned, as counsel for the United States of America and Jeffrey Epstein, jointly write to you to provide information relevant to your service as a Special Master in the selection of an attorney to represent several young women who may have civil damages claims against Mr. Epstein. The U.S. Attorney's Office and the Federal Bureau of Investigation (jointly referred to as the "United States") have conducted an investigation of Jeffrey Epstein regarding his solicitation of minor females in Palm Beach County to engage in prostitution. Mr. Epstein, through his assistants, would recruit underage females to travel to his home in Palm Beach to engage in lewd conduct in exchange for money. Based upon the investigation, the United States has identified forty (40) young women who can be characterized as victims pursuant to 18 USC 2255. Some of those women went to Mr. Epstein's home only once, some went there as much as 100 times or more. Some of the women's conduct was limited to performing a topless or nude massage while Mr. Epstein masturbated himself. For other women, the conduct escalated to full sexual intercourse. As part of the resolution of the case, Epstein has agreed that he would not contest jurisdiction in the Southern District of Florida for any victim who chose to sue him for damages pursuant to 18 USC 2255. Mr. Epstein agreed to provide an attorney for victims who elected to proceed exclusively pursuant to that section, and agreed to waive any challenge to liability under that section up to an amount agreed to by the parties. The parties have agreed to submit the selection of an attorney to a Special Master...."

7. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that L.M. was, in fact, a victim of Epstein's criminal abuse because L.M. was one of the

minor females that the United States Attorney's Office recognized as a victim. L.M.'s sworn deposition testimony and the adverse inference drawn from Epstein's refusal to testify confirm that Epstein began sexually assaulting L.M. when she was 13 years old and continued to molest her on more than fifty (50) occasions over three (3) years. Epstein Depo., Attachment #1, at 17 ("Q: Did you . . . ever engage in any sexual conduct with L.M.?" A: [Invocation of the Fifth Amendment].); *see also* Epstein Depo., April 14, 2010, Attachment #6, at 456 ("Q: LM was an underage female that you first abused when she was 13 years old; is that correct?" A: [Invocation of Fifth Amendment].) .

8. Epstein was also given ample opportunity to explain why he engaged in sexual activity with L.M. beginning when L.M. was 13 years old and why he has molested minors on an everyday basis for years, and he invoked his 5th amendment right rather than provide explanation. *See* Epstein Deposition, February 17, 2010, at 11-12, 30-31 (Deposition Attachment # 7).

9. Epstein also sexually assaulted E.W., beginning when she was 14 years old and did so on numerous occasions. *See* E.W. Depo., Attachment #4 at 215-216.

10. Another of the minor girls Epstein sexually assaulted was Jane Doe; the abuse began when Jane Doe was 14 years old. Rather than incriminate himself, Epstein invoked the 5th amendment to questions about him digitally penetrating Doe's vagina, using vibrators on her vagina and masturbating and ejaculating in her presence. Epstein Depo., April 14, 2010, Attachment #6, at 420, 464, 468.

11. When Edwards's clients L.M., E.W., and Jane Doe were 13 or 14 years old, each was brought to Epstein's home multiple times by another underage victim. Epstein engaged in

one or more of the following acts with each of the then-minor girls at his mansion: receiving a topless or completely nude massage; using a vibrator on her vagina; masturbating in her presence; ejaculating in her presence; touching her breast or buttocks or vagina or the clothes covering her sexual organs; and demanding that she bring him other underage girls. Epstein and his co-conspirators used the telephone to contact these girls to entice or induce them into going to his mansion for sexual abuse. Epstein also made E.W. perform oral sex on him and was to perform sex acts on Nadia Marcinkova (Epstein's live-in sex slave) in Epstein's presence. *See* Plaintiff Jane Doe's Notice Regarding Evidence of Similar Acts of Sexual Assault, filed in Jane Doe v. Epstein, No. 08-cv-80893 (S.D. Fla. 2010), as DE 197, (hereinafter "Rule 413 Notice") (Exhibit "E"); Jane Doe Depo., Attachment #2, at 379-380; L.M. Depo., Attachment #3, at 416; E.W. Depo, Attachment #4, at 205.

12. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that yet another of the minor girls Epstein sexually assaulted was C.L. When she was approximately 15 years old, C.L. was brought to Epstein's home by another underage victim. While a minor, she was at Epstein's home on multiple occasions. Epstein engaged in one or more of the following acts with her while she was a minor at his house - topless or completely nude massage on Epstein; Epstein used a vibrator on her vagina; Epstein masturbated in her presence; Epstein ejaculated in her presence; Epstein also demanded that she bring him other underage girls. *See* Rule 413 Notice, Exhibit "E"; Incident Report, Exhibit "A."

13. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that yet another girl Epstein sexually assault was A.H. When she was approximately 16 years old, she was brought to Epstein's home by another underage victim. While a minor, she

was at Epstein's home on multiple occasions. Epstein engaged in one or more of the following acts with her while she was a minor at his house - topless or completely nude massage on Epstein; Epstein used a vibrator on her vagina; Epstein masturbated in her presence; Epstein ejaculated in her presence; Epstein touched her breast or buttock or vagina or the clothes covering her sexual organs; was made to perform sex acts on Epstein; made to perform sex acts on Nadia Marcinkova in Epstein's presence. Epstein also forcibly raped this underage victim, as he held her head down against her will and pumped his penis inside her while she was screaming "No". *See* Rule 413 Notice, Exhibit "E"; Incident Report, Exhibit "A", at 41 (specifically discussing the rape):

> "[A.H.] remembered that she climaxed and was removing herself from the massage table. [A.H.] asked for a sheet of paper and drew the massage table in the master bathroom and where Epstein, Marcinkova and she were. Epstein turned [A.H.] on to her stomach on the massage bed and inserted his penis into her vagina. [A.H.] stated Epstein began to pump his penis in her vagina. [A.H.] became upset over this. She said her head was being held against the bed forcibly, as he continued to pump inside her. She screamed no, and Epstein stopped ...."
>
> "[A.H.] advised there were times that she was so sore when she left Epstein's house. [A.H.] advised she was ripped, torn, in her vagina area. [A.H.] advised she had difficulty walking to the car after leaving the house because she was so sore."

14. Without detailing each fact known about Epstein's abuse of the many underage girls, Edwards has had a good faith basis to believe and did in fact believe at all relevant times that Epstein also abused other victims in ways closely similar to those described in the preceding paragraphs. Epstein's additional victims include the following (among many other) young girls: S.G.; A.D.; V.A.; N.R.; J.S.; V.Z.; J.A.; F.E.; M.L.; M.D.; D.D.; and D.N. These girls were between the ages of 13 and 17 when Epstein abused them. *See* Rule 413 Notice, Exhibit E; Deposition of E.W., Deposition Attachment #4.

15. One of Mr. Epstein's household employees, Mr. Alfredo Rodriguez, saw numerous underage girls coming into Epstein's mansion for purported "massages." *See* Rodriguez Depo. at 242-44 (Deposition Attachment #8). Rodriguez was aware that "sex toys" and vibrators were found in Epstein's bedroom after the purported massages. *Id.* at 223-28. Rodriguez thought what Epstein was doing was wrong, given the extreme youth of the girls he saw. *Id.* at 230-31..

16. Alfredo Rodriguez took a journal from Epstein's computer that reflected many of the names of underage females Epstein abused across the country and the world, including locations such as Michigan, California, West Palm Beach, New York, New Mexico, and Paris, France. *See* Journal (hereinafter "The Journal" or "Holy Grail") (Exhibit "F") (identifying, among other Epstein acquaintances, females that Rodriguez believes were underage under the heading labeled "Massages").

17. Rodriguez was later charged in a criminal complaint with obstruction of justice in connection with trying to obtain $50,000 from civil attorneys pursuing civil sexual assault cases against Epstein as payment for producing the book to the attorneys. *See* Criminal Complaint at 2, U.S. v. Rodriguez, No. 9:10-CR-80015-KAM (S.D. Fla. 2010) (Exhibit "G"). Rodriguez stated he needed money because the journal was his "property" and that he was afraid that Jeffrey Epstein would make him "disappear" unless he had an "insurance policy" (i.e., the journal). *Id.* at 3. Because of the importance of the information in the journal to the civil cases, Mr. Rodriguez called it "The Holy Grail."

18. In the "Holy Grail" or "The Journal," among the many names listed (along with the abused girls) are some of the people that Epstein alleges in his Complaint had "no connection

whatsoever" with the litigation in this case. *See, e.g.,* Journal, Exhibit F, at 85 (Donald Trump); at 9 (Bill Clinton phone numbers listed under "Doug Bands").

Federal Investigation and Plea Agreement With Epstein

19. In approximately 2005, the FBI and the U.S. Attorney's Office in the Southern District of Florida learned of Epstein's repeated sexual abuse of minor girls. They began a criminal investigation into federal offenses related to his crimes. *See* U.S. Attorney's Correspondence, Exhibit "C".

20. At all relevant times Edwards has had a good faith basis to believe and did in fact believe that to avoid the Government learning about his abuse of minor girls, Epstein threatened his employees and demanded that they not cooperate with the government. Epstein's aggressive witness tampering was so severe that the United States Attorney's Office prepared negotiated plea agreements containing these charges. For example, in a September 18, 2007, email from AUSA Villafaña to Lefkowitz (attached hereto as Exhibit "H"), she attached the proposed plea agreement describing Epstein's witness tampering as follows:

> "UNITED STATES vs. JEFFREY EPSTEIN PLEA PROFFER"
>
> On August 21, 2007, FBI Special Agents E. Nesbitt Kuyrkendall and Jason Richards traveled to the home of Leslie Groff to serve her with a federal grand jury subpoena with an investigation pending in the Southern District of Florida. Ms. Groff works as the personal assistant of the defendant. Ms. Groff began speaking with the agents and then excused herself to go upstairs to check on her sleeping child. While upstairs, Ms. Groff telephoned the defendant, Jeffrey Epstein, and informed him that the FBI agents were at her home. Mr. Epstein instructed Ms. Groff not to speak with the agents and reprimanded her for allowing them into her home. Mr. Epstein applied pressure to keep Ms. Groff from complying with the grand jury subpoenas that the agents had served upon her. In particular, Mr. Epstein warned Ms. Groff against turning over documents and electronic evidence responsive to the subpoena and pressured her to delay her

> appearance before the grand jury in the Southern District of Florida. This conversation occurred when Mr. Epstein was aboard his privately owned civilian aircraft in Miami in the Southern District of Florida. His pilot had filed a flight plan showing the parties were about to return to Teterboro, NJ. After the conversation with Ms. Groff, Mr. Epstein became concerned that the FBI would try to serve his traveling companion, Nadia Marcinkova, with a similar grand jury subpoena. In fact, the agents were preparing to serve Ms. Marcinkova with a target letter when the flight landed in Teterboro. Mr. Epstein then redirected his airplane, making the pilot file a new flight plan to travel to the US Virgin Islands instead of the New York City area, thereby keeping the Special Agents from serving the target letter on Nadia Marcinkova. During the flight, the defendant verbally harassed Ms. Marcinkova, harassing and pressuring her not to cooperate with the grand jury's investigation, thereby hindering and dissuading her from reporting the commission of a violation of federal law to a law enforcement officer, namely, Special Agents of the FBI. Epstein also threatened and harassed Sarah Kellen against cooperating against him as well.

21. Edwards learned that the Palm Beach police department investigation ultimately led to the execution of a search warrant at Epstein's mansion in October 2005. *See* Police Incident Report, Exhibit "A".

22. Edwards learned that at around the same time, the Palm Beach Police Department also began investigating Epstein's sexual abuse of minor girls. They also collected evidence of Epstein's involvement with minor girls and his obsession with training sex slaves, including pulling information from Epstein's trash. Their investigation showed that Epstein ordered from Amazon.com on about September 4, 2005, such books as: SM101: A Realistic Introduction, by Jay Wiseman; SlaveCraft: Roadmaps for Erotic Servitude - Principles, Skills, and Tools, by Guy Baldwin; and Training with Miss Abernathy: A Workbook for Erotic Slaves and Their Owners, by Christina Abernathy. *See* Receipt for Sex Slave Books (Exhibit "I").

23. The Palm Beach incident reports provided Edwards with the names of numerous witnesses that participated in Epstein's child molestation criminal enterprise and also provided

Edwards with some insight into how far-reaching Epstein's power was and how addicted Epstein was to sex with children. *See* Incident Report, Exhibit "A".

24. The Palm Beach Police Department also collected Epstein's message pads, which provided other names of people that also knew Epstein's scheme to molest children. *See* Message Pads (Exhibit "J") (note: the names of underage females have been redacted to protect the anonymity of the underage sex abuse victims). Those message pads show clear indication that Epstein's staff was frequently working to schedule multiple young girls between the ages of 12 and 16 years old literally every day, often two or three times per day. *Id.*

25. In light of all of the information of numerous crimes committed by Epstein, Edwards learned that the U.S. Attorney's Office began preparing the filing of federal criminal charges against Epstein. For example, in addition to the witness tampering and money laundering charges the U.S. Attorney's Office prepared an 82-page prosecution memo and a 53-page indictment of Epstein related to his sexual abuse of children. On September 19, 2007, at 12:14 PM, AUSA Villafaña wrote to Epstein's counsel, Jay Lefkowitz, "Jay - I hate to have to be firm about this, but we need to wrap this up by Monday. I will not miss my indictment date when this has dragged on for several weeks already and then, if things fall apart, be left in a less advantageous position than before the negotiations. I have had an 82-page pros memo and 53-page indictment sitting on the shelf since May to engage in these negotiations. There has to be an ending date, and that date is Monday." These and other communications are within the correspondence attached as Composite Exhibit "C."

26. Edwards learned that rather than face the filing of federal felony criminal charges, Epstein (through his attorneys) engaged in plea bargain discussions. As a result of those

discussions, on September 24, 2007, Epstein signed an agreement with the U.S. Attorney's Office for the Southern District of Florida. Under the agreement, Epstein agreed to plead guilty to an indictment pending against him in the 15th Judicial Circuit in and for Palm Beach County charging him with solicitation of prostitution and procurement of minors for prostitution. Epstein also agreed that he would receive a thirty month sentence, including 18 months of jail time and 12 months of community control. In exchange, the U.S. Attorney's Office agreed not to pursue any federal charges against Epstein. *See* Non-Prosecution Agreement (Exhibit "K").

27. Part of the Non-Prosecution Agreement that Epstein negotiated was a provision in which the federal government agreed not to prosecute Epstein's co-conspirators. The co-conspirators procured minor females to be molested by Epstein. One of the co-conspirators - Nadia Marcinkova -even participated in the sex acts with minors (including E.W.) and Epstein. *See* Incident Report, Exhibit "A", at 40-42, 49-51; Deposition of Nadia Marcinkova, April 13, 2010, (hereinafter "Marcinkova Depo.") at 11 (Deposition attachment #9).

28. Under the Non-Prosecution Agreement, Epstein was to use his "best efforts" to enter into his guilty pleas by October 26, 2007. However, Edwards learned that Epstein violated his agreement with the U.S. Attorney's Office to do so and delayed entry of his plea. *See* Letter from U.S. Attorney R. Alexander Acosta to Lilly Ann Sanchez, Dec. 19, 2007 (Exhibit "L").

29. On January 10, 2008 and again on May 30, 2008 E.W. and L.M. received letters from the FBI advising them that "[t]his case is currently under investigation. This can be a lengthy process and we request your continued patience while we conduct a thorough investigation." Letters attached at Composite Exhibit "M". This document is evidence that the FBI did not notify E.W. and L.M. that a plea agreement had already been reached that would

block federal prosecution of Epstein. Nor did the FBI notify E.W. and L.M. of any of the parts of the plea agreement. Nor did the FBI or other federal authorities confer with E.W. and L.M. about the plea. *See id.*

30. In 2008, Edwards believed in good faith that criminal prosecution of Epstein was extremely important to his clients E.W. and L.M. and that they desired to be consulted by the FBI and/or other representatives of the federal government about the prosecution of Epstein. The letters that they had received around January 10, 2008, suggested that a criminal investigation of Epstein was on-going and that they would be contacted before the federal government reached any final resolution of that investigation. *See id.*

Edwards Agrees to Serve as Legal Counsel for Three Victims of Epstein's Sexual Assaults

31. In about April 2008, Bradley J. Edwards, Esq., was a licensed attorney in Florida, practicing as a sole practitioner. As a former prosecutor, he was well versed in civil cases that involved criminal acts, including sexual assaults. Three of the many girls Epstein had abused – L.M., E.W., and Jane Doe – all requested that Edwards represent them civilly and secure appropriate monetary damages against Epstein for repeated acts of sexual abuse while they were minor girls. Two of the girls (L.M. and E.W.) also requested that Edwards represent them in connection with a concern that the Federal Bureau of Investigation (FBI) and U.S. Attorney's Office might be arranging a plea bargain for the criminal offenses committed by Epstein without providing them the legal rights to which they were entitled (including the right to be notified of plea discussions and the right to confer with prosecutors about any plea arrangement). *See*

Affidavit of Bradley J. Edwards, Esq. at ¶1 - 2, ¶4 (hereinafter "Edwards Affidavit") (Exhibit "N").

32. On June 13, 2008, attorney Edwards agreed to represent E.W.; on July 2, 2008, attorney Edwards agreed to represent Jane Doe; and, on July 7, 2008, attorney Edwards agreed to represent L.M. in connection with the sexual assaults committed by Epstein and to insure that their rights as victims of crimes were protected in the criminal process on-going against Epstein. Mr. Edwards and his three clients executed written retention agreements. *See id.* at ¶2.

33. In mid June of 2008, Edwards contacted AUSA Villafaña to inform her that he represented Jane Doe #1 and, later, Jane Doe #2. AUSA Villafaña did not advise that a plea agreement had already been negotiated with Epstein's attorneys that would block federal prosecution. To the contrary, AUSA Villafaña mentioned a possible indictment. AUSA Villafaña did indicate that federal investigators had concrete evidence and information that Epstein had sexually molested many underage minor females, including E.W., LM, and Jane Doe. *See id.* at ¶4.

34. Edwards also requested from the U.S. Attorney's Office the information that they had collected regarding Epstein's sexual abuse of his clients. However, the U.S. Attorney's Office, declined to provide any such information to Edwards. It similarly declined to provide any such information to the other attorneys who represented victims of Epstein's sexual assaults. At the very least, this includes the items that were confiscated in the search warrant of Epstein's home, including dildos, vibrators, massage table, oils, and additional message pads. *See* Property Receipt (Exhibit "O").

35. On Friday, June 27, 2008, at approximately 4:15 p.m., AUSA Villafaña received a copy of Epstein's proposed state plea agreement and learned that the plea was scheduled for 8:30 a.m., Monday, June 30, 2008. AUSA Villafaña called Edwards to provide notice to his clients regarding the hearing. AUSA Villafaña did not tell Attorney Edwards that the guilty pleas in state court would bring an end to the possibility of federal prosecution pursuant to the plea agreement. *See* Edwards Affidavit, Exhibit "N", at ¶6.

36. Under the Crime Victims' Rights Act (CVRA), 18 U.S.C. § 3771, victims of federal crimes – including E.W. and L.M. – are entitled to basic rights during any plea bargaining process; including the right to be treated with fairness, the right to confer with prosecutors regarding any plea, and the right to be heard regarding any plea. The process that was followed leading to the non-prosecution of Epstein violated these rights of E.W. and L.M. *See* Emergency Petn. for Victim's Enforcement of Crime Victim's Rights, No. 9:08-CV-80736-KAM (S.D. Fla. 2008) (Exhibit "P").

37. Because of the violation of the CVRA, on July 7, 2008, Edwards filed an action in the U.S. District Court for the Southern District of Florida, Case No. 9:08-CV-80736, seeking to enforce the rights of E.W. and L.M. That action alleged that the U.S. Attorney's Office had failed to provide E.W. and L.M. the rights to which they were entitled under the Act, including the right to be notified about a plea agreement and to confer with prosecutors regarding it. *See id.*

38. On July 11, 2008, Edwards took E.W. and L.M. with him to the hearing on the CVRA action. It was only at this hearing that both victims learned for the first time that the plea deal was already done with Epstein and that the criminal case against Epstein had been

effectively terminated by the U.S. Attorney's office. *See* Hearing Transcript, July 11, 2008 (Exhibit "Q").

39. Edwards learned that Jane Doe felt so strongly that the plea bargain was inappropriate that she made her own determination to appear on a television program and exercise her First Amendment rights to criticize the unduly lenient plea bargain Epstein received in a criminal case.

40. The CVRA action that Edwards filed was recently administratively closed and Edwards filed a Motion to reopen that proceeding. *See* No. 9:08-CV-80736 (S.D. Fla.).

Epstein's Entry of Guilty Pleas to Sex Offenses

41. Ultimately, on June 30, 2008, in the Fifteenth Judicial Circuit in Palm Beach County, Florida, defendant Epstein, entered pleas of "guilty" to various Florida state crimes involving the solicitation of minors for prostitution and the procurement of minors for the purposes of prostitution. *See* Plea Colloquy (Exhibit "R").

42. As a condition of that plea, and in exchange for the Federal Government not prosecuting the Defendant, Epstein additionally entered into an agreement with the Federal Government acknowledging that approximately thirty-four (34) other young girls could receive payments from him under the federal statute providing for compensation to victims of child sexual abuse, 18 U.S.C. § 2255. As had been agreed months before, the U.S. Attorney's Office did not prosecute Epstein federally for his sexual abuse of these minor girls. *See* Addendum to Non-Prosecution Agreement (Exhibit "S") (in redacted form to protect the identities of the minors involved).

43. Because Epstein became a convicted sex offender, he was not to have contact with any of his victims. During the course of his guilty pleas on June 30, 2008, Palm Beach Circuit Court Judge Deborah Dale Pucillo ordered Epstein "not to have any contact, direct or indirect" with any victims. She also expressly stated that her no-contact order applied to "all of the victims." Similar orders were entered by the federal court handling some of the civil cases against Epstein. The federal court stated that it "finds it necessary to state clearly that Defendant is under this court's order not to have direct *or indirect contact* with any plaintiffs . . . ." Order, Case No. 9:08-cv-80119 (S.D. Fla. 2008), [DE 238] at 4-5 (emphasis added); *see also* Order, Case No. 9:08-cv-80893, [DE 193] at 2 (emphasis added).

Edwards Files Civil Suits Against Epstein

44. Edwards had a good faith belief that his clients felt angry and betrayed by the criminal system and wished to prosecute and punish Epstein for his crimes against them in whatever avenue remained open to them. On August 12, 2008, at the request of his client Jane Doe, Brad Edwards filed a civil suit against Jeffrey Epstein to recover damages for his sexual assault of Jane Doe. *See* Edwards Affidavit, "N" at ¶7. Included in this complaint was a RICO count that explained how Epstein ran a criminal conspiracy to procure young girls for him to sexually abuse. *See* Complaint, Jane Doe v. Epstein (Exhibit "T").

45. On September 11, 2008, at the request of his client E.W., Brad Edwards filed a civil suit against Jeffrey Epstein to recover damages for his sexual assault of E.W. *See* Complaint, E.W. v. Epstein (Exhibit "U").

46. On September 11, 2008, at the request of his client L.M.., Brad Edwards filed a civil suit against Jeffrey Epstein to recover damages for his sexual assault of L.M. *See* Complaint, L.M. v. Epstein, (Exhibit "V").

47. Jane Doe's federal complaint indicated that she sought damages of more than $50,000,000. Listing the amount of damages sought in the complaint was in accord with other civil suits that were filed against Epstein (before any lawsuit filed by Edwards). *See* Complaint, Jane Doe #4 v. Epstein (Exhibit "W") (filed by Herman and Mermelstein, PA).

48. At about the same time as Edwards filed his three lawsuits against Epstein, other civil attorneys were filing similar lawsuits against Epstein. For example, on or about April 14, 2008 another law firm, Herman and Mermelstein, filed the first civil action against Epstein on behalf of one of its seven clients who were molested by Epstein. The complaints that attorney Herman filed on behalf of his seven clients were similar in tenor and tone to the complaint that Edwards filed on behalf of his three clients. *See id.*

49. Over the next year and a half, more than 20 other similar civil actions were filed by various attorneys against Epstein alleging sexual assault of minor girls. These complaints were also similar in tenor and tone to the complaint that Edwards filed on behalf of his clients. These complaints are all public record and have not been attached, but are available in this Court's files and the files of the U.S. District Court for the Southern District of Florida.

50. In addition to the complaints filed against Epstein in Florida, a female in New York, Ava Cordero, filed a lawsuit against Epstein in New York making similar allegations - that Epstein paid her for a massage then forced her to give him oral sex and molested her in other ways when she was only 16 years old. Cordero was born a male, and in her complaint she

alleges that Epstein told her during the "massage", "I love how young you are. You have a tight butt like a baby". *See* Jeff Epstein Sued for "Repeated Sexual Assaults" on Teen, New York Post, October 17, 2007, by Dareh Gregorian, link at: http://www.nypost.com/p/news/regional/item_44zlWyLUFH7R1OUtKYGPbP;jsessionid=6CA3EBF1BEF68F5DE14BFB2CAA5C37E0. *See* Article attached hereto as Exhibit "X".

51. Edwards's three complaints against Epstein contained less detail about sexual abuse than (as one example) a complaint filed by attorney Robert Josephsberg from the law firm of Podhurst Orseck. *See* Complaint, Jane Doe 102 v. Epstein (Exhibit "B"). As recounted in detail in this Complaint, Jane Doe 102 was 15 years old when Ghislaine Maxwell discovered her and lured her to Epstein's house. Maxwell and Epstein forced her to have sex with both of them and within weeks Maxwell and Epstein were flying her all over the world. According to the Complaint, Jane Doe 102 was forced to live as one of Epstein's underage sex slaves for years and was forced to have sex with not only Maxwell and Epstein but also other politicians, businessmen, royalty, academicians, etc. She was even made to watch Epstein have sex with three 12-year-old French girls that were sent to him for his birthday by a French citizen that is a friend of Epstein's. Luckily, Jane Doe 102 escaped to Australia to get away from Epstein and Maxwell's sexual abuse.

52. Edwards learned that in addition to civil suits that were filed in court against Epstein, at around the same time other attorneys engaged in pre-filing settlement discussions with Epstein. Rather than face filed civil suits in these cases, Epstein paid money settlements to more than 15 other women who had sexually abused while they were minors. *See* articles regarding settlements attached hereto as Composite Exhibit "Y."

Epstein's Obstruction of Normal Discovery and Attacks on His Victims

53. Once Edwards filed his civil complaints for his three clients, he began the normal process of discovery for cases such as these. He sent standard discovery requests to Epstein about his sexual abuse of the minor girls, including requests for admissions, request for production, and interrogatories. *See* Edwards Affidavit, Exhibit "N", at ¶¶11-19 and 25.

Rather than answer any substantive questions about his sexual abuse and his conspiracy for procuring minor girls for him to abuse, Epstein invoked his 5th amendment right against self-incrimination. An example of Epstein's refusal to answer is attached as Composite Exhibit "Z" (original discovery propounded to Epstein and his responses invoking 5th amendment).

54. During the discovery phase of the civil cases filed against Epstein, Epstein's deposition was taken at least five times. During all of those depositions, Epstein refused to answer any substantive questions about his sexual abuse of minor girls. *See, e.g.,* Deposition Attachments 1, 6 and 7.

55. During these depositions, Epstein further attempted to obstruct legitimate questioning by inserting a variety of irrelevant information about his case. As one of innumerable examples, on March 8, 2010, Mr. Horowitz, representing seven victims, Jane Doe's 2-8, asked, "Q: In 2004, did you rub Jane Doe 3's vagina? A: Excuse me. I'd like to answer that question, as I would like to answer mostly every question you've asked me here today; however, upon advice of counsel, I cannot answer that question. They've advised me I must assert my Sixth Amendment, Fifth Amendment and Fourteenth Amendment Rights against self--excuse me, against--under the Constitution. And though your partner, Jeffrey Herman, was disbarred after filing this lawsuit [a statement that was untrue], Mr. Edwards' partner sits in jail for

fabricating cases of a sexual nature fleecing unsuspecting Florida investors and others out of millions of dollars for cases of a sexual nature with--I'd like to answer your questions; however if I--I'm told that if I do so, I risk losing my counsel's representation; therefore I must accept their advice." Epstein deposition, March 8, 2010, at 106 (Deposition attachment #10).

56. When Edwards had the opportunity to take Epstein's deposition, he only asked reasonable questions, all of which related to the merits of the cases against Epstein. All depositions of Epstein in which Mr. Edwards participated on behalf of his clients are attached to this motion. *See* Edwards Affidavit, Exhibit "N" at ¶11 and Deposition attachments #1, 6, 7, 10, 11, 12, and 13. Cf. with Deposition of Epstein taken by an attorney representing BB (one in which Edwards was not participating), http://www.youtube.com/watch?v=V-dqoEyYXx4; and http://www.youtube.com/watch?v=YCNiYltW-r0

57. Edwards's efforts to obtain information about Epstein's organization for procuring young girls was also blocked because Epstein's co-conspirators took the Fifth. Deposition of Sarah Kellen, March 24, 2010 (hereinafter "Kellen Depo.") (Deposition attachment #14); Deposition of Nadia Marcinkova, April 13, 2010, (Deposition attachment #9); Deposition of Adriana Mucinska Ross, March 15, 2010 (hereinafter "Ross Depo.") (Deposition attachment #15). Each of these co-conspirators invoked their respective rights against self-incrimination as to all relevant questions, and the depositions have been attached.

58. At all relevant times Edwards has had a good faith basis to believe and did in fact believe Sarah Kellen was an employee of Epstein's and had been identified as a defendant in at least one of the complaints against Epstein for her role in bringing girls to Epstein's mansion to be abused. At the deposition, she was represented by Bruce Reinhart. She invoked the Fifth on

all substantive questions regarding her role in arranging for minor girls to come to Epstein's mansion to be sexually abused. Reinhart had previously been an Assistant United States Attorney in the U.S. Attorney's Office for the Southern District of Florida when Epstein was being investigated criminally by Reinhart's office. Reinhart left the United States Attorney's Office and was immediately hired by Epstein to represent Epstein's pilots and certain co-conspirators during the civil cases against Epstein. *See* Edwards Affidavit, Exhibit "N" at ¶11.

59. Edwards also had other lines of legitimate discovery blocked through the efforts of Epstein and others. For example, Edwards learned through deposition that Ghislaine Maxwell was involved in managing Epstein's affairs and companies. *See* deposition of Epstein's house manager Janusz Banziak, February 16, 2010 at page 14, lines 20-23 (Deposition Attachment #16); *See* deposition of Epstein's housekeeper Louella Rabuyo, October 20, 2009, page 9, lines 17-25 (Deposition Attachment #17); *See* deposition of Epstein's pilot Larry Eugene Morrison, October 6, 2009, page 102-103 (Deposition Attachment #18); *See* deposition of Alfredo Rodriguez, August 7, 2009, page 302-306 and 348 (Deposition Attachment #8); *See* also Prince Andrew's Friend, Ghislaine Maxwell, Some Underage Girls and A Very Disturbing Story, September 23, 2007 by Wendy Leigh, link at http://www.redicecreations.com/article.php?id=1895OHANNA SJOBERG. Exhibit "AA".

60. Alfredo Rodriguez testified that Maxwell took photos of girls without the girls' knowledge, kept the images on her computer, knew the names of the underage girls and their respective phone numbers and other underage victims were molested by Epstein and Maxwell together. *See* Deposition of Rodriguez, Deposition attachment # 8 at 64, 169-170 and 236.

61. In reasonable reliance on this and other information, Edwards served Maxwell for deposition in 2009. *See* Deposition Notice attached as Exhibit "BB." Maxwell was represented by Brett Jaffe of the New York firm of Cohen and Gresser, and Edwards understood that her attorney was paid for (directly or indirectly) by Epstein. She was reluctant to give her deposition, and Edwards tried to work with her attorney to take her deposition on terms that would be acceptable to both sides. The result was the attached confidentiality agreement, under which Maxwell agreed to drop any objections to the deposition, attached hereto as Exhibit "CC." Maxwell, however, contrived to avoid the deposition. On June 29, 2010, one day before Edwards was to fly to NY to take Maxwell's deposition, her attorney informed Edwards that Maxwell's mother was deathly ill and Maxwell was consequently flying to England with no intention of returning to the United States. Despite that assertion, Ghislaine Maxwell was in fact in the country on July 31, 2010, as she attended the wedding of Chelsea Clinton (former President Clinton's daughter) and was captured in a photograph taken for OK magazine. Photos from Issue 809 of the publication *See* US Weekly dated August 16, 2010 are attached hereto as Exhibit "DD" and Edwards Affidavit, Exhibit "N" at ¶12.

62. Maxwell is not the only important witness to lie to avoid deposition by Edwards. Upon review of the message pads that were taken from Epstein's home in the police trash pulls, *see* Exhibit "J" *supra*, many were from Jean Luc Brunel, a French citizen and one of Epstein's closest pals. He left messages for Epstein. One dated 4/1/05 said, "He has a teacher for you to teach you how to speak Russian. She is 2x8 years old, not blonde. Lessons are free and you can have your 1st today if you call." *See* Messages taken from Jean Luc Brunel are attached hereto as Exhibit "EE." In light of these circumstances of the case, this message reasonably suggested to

Edwards that Brunel might have been procuring two eight-year-old girls for Epstein to sexually abuse. According to widely circulated press reports reviewed by Edwards, Brunel is in his sixties and has a reputation throughout the world (and especially in the modeling industry) as a cocaine addict that has for years molested children through modeling agencies while acting as their agent – conduct that has been the subject of critical reports, books, several news articles, and a 60 Minutes documentary on Brunel's sexual exploitation of underage models. *See* http://bradmillershero.blogspot.com/2010/08/women-are-objects.html, attached hereto as Exhibit "FF."

63. Edwards learned that Brunel is also someone that visited Epstein on approximately 67 occasions while Epstein was in jail. *See* Epstein's jail visitor log attached as Exhibit "GG."

64. Edwards learned that Brunel currently runs the modeling agency MC2, a company for which Epstein provides financial support. *See* Message Pad's attached as Exhibit "J" *supra* and Sworn Statement of MC2 employee Maritza Vasquez, June 15, 2010, "Maritza Vasquez Sworn Statement" attached at Exhibit "HH" at 1-16.

65. Employees of MC2 told Edwards that Epstein's numerous condos at 301 East 66 Street in New York were used to house young models. Edwards was told that MC2 modeling agency, affiliated with Epstein and Brunel brought underage girls from all over the world, promising them modeling contracts. Epstein and Brunel would then obtain a visa for these girls, then would charge the underage girls rent, presumably to live as underage prostitutes in the condos. *See* Maritza Vasquez Sworn Statement, Exhibit "HH" at 7-10, 12-15, 29-30, 39-41, 59-60 and 62-67.

66. In view of this information suggesting Brunel could provide significant evidence of Epstein's trafficking in young girls for sexual abuse, Edwards had Brunel served in New York for deposition. *See* Notice of Deposition of Jean Luc Brunel attached hereto as Exhibit "II." Before the deposition took place, Brunel's attorney (Tama Kudman of West Palm Beach) contacted Edwards to delay the deposition date. Eventually Kudman informed Edwards in January 2009 that Brunel had left the country and was back in France with no plans to return. This information was untrue; Brunel was actually staying with Epstein in West Palm Beach. *See* Banasiak deposition, deposition attachment #16 at 154-160 and 172-175; see also pages from Epstein's probation file evidencing Jean Luc Brunel (JLB) staying at his house during that relevant period of time attached Exhibit "JJ". As a result, Edwards filed a Motion for Contempt, attached hereto as Exhibit "KK" (Because Epstein settled this case, the motion was never ruled upon.)

67. Edwards was also informed that Epstein paid for not only Brunel's representation during the civil process but also paid for legal representation for Sarah Kellen (Epstein's executive assistant and procurer of girls for him to abuse), Larry Visoski (Epstein's personal pilot), Dave Rogers (Epstein's personal pilot), Larry Harrison (Epstein's personal pilot), Louella Rabuyo (Epstein's housekeeper), Nadia Marcinkova (Epstein's live-in sex slave), Ghislaine Maxwell (manager of Epstein's affairs and businesses), Mark Epstein (Epstein's brother), and Janusz Banasiak (Epstein's house manager) It was nearly impossible to take a deposition of someone that would have helpful information that was not represented by an attorney paid for by Epstein. *See* Edwards Affidavit, Exhibit "N" at ¶11.

68. While Epstein and others were preventing any legitimate discovery into his sexual abuse of minor girls, at the same time he was engaging (through his attorneys) in brutal questioning of the girls who had filed civil suits against him, questioning so savage that it made local headlines. *See* Jane Musgrave, *Victims Seeking Sex offender's Millions See Painful Pasts Used Against Them,* Palm Beach Post News, Jan. 23, 2010, available at http://www.palmbeachpost.com/news/crime/victims-seeking-sex-offenders-millions-see-painful-pasts-192988.html attached hereto as Exhibit "LL."

Edwards Pursues Other Lines of Discovery

69. Because of Epstein's thwarting of discovery and attacks on Edwards's clients, Edwards was forced to pursue other avenues of discovery. Edwards only pursued legitimate discovery designed to further the cases filed against Epstein. *See* Edwards Affidavit, Exhibit "N" at ¶11.

70. Edwards notified Epstein's attorneys of his intent to take Bill Clinton's deposition. Edwards possessed a legitimate basis for doing so: (a) Clinton was friends with Ghislaine Maxwell who was Epstein's longtime companion and helped to run Epstein's companies, kept images of naked underage children on her computer, helped to recruit underage children for Epstein, engaged in lesbian sex with underage females that she procured for Epstein, and photographed underage females in sexually explicit poses and kept child pornography on her computer; (b) it was national news when Clinton traveled with Epstein aboard Epstein's private plane to Africa and the news articles classified Clinton as Epstein's friend. (c) the complaint filed on behalf of Jane Doe No. 102 stated generally that she was required by Epstein to be sexually exploited by not only Epstein but also Epstein's "adult male peers, including royalty,

politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories Clinton and acquaintances of Clinton fall into. The flight logs showed Clinton traveling on Epstein's plane on numerous occasions between 2002 and 2005. *See* Flight logs attached hereto as Exhibit "MM." Clinton traveled on many of those flights with Ghislaine Maxwell, Sarah Kellen, and Adriana Mucinska, - all employees and/or co-conspirators of Epstein's that were closely connected to Epstein's child exploitation and sexual abuse. The documents clearly show that Clinton frequently flew with Epstein aboard his plane, then suddenly stopped - raising the suspicion that the friendship abruptly ended, perhaps because of events related to Epstein's sexual abuse of children. Epstein's personal phone directory from his computer contains e-mail addresses for Clinton along with 21 phone numbers for him, including those for his assistant (Doug Band), his schedulers, and what appear to be Clinton's personal numbers. This information certainly leads one to believe that Clinton might well be a source of relevant information and efforts to obtain discovery from him were reasonably calculated to lead to admissible evidence. *See* Exhibits "B", "F" "AA", "DD", and "MM" and Edwards Affidavit, Exhibit "N" at ¶15.

71. Bradley J. Edwards, Esq., provided notice that he intended to take the deposition of Donald Trump. Edwards possessed a legitimate basis for doing so: (a) The message pads confiscated from Epstein's home indicated that Trump called Epstein's West Palm Beach mansion on several occasions during the time period most relevant to my Edwards's clients' complaints; (b) Trump was quoted in a *Vanity Fair* article about Epstein as saying "I've known Jeff for fifteen years. Terrific guy," "He's a lot of fun to be with. It is even said that he likes beautiful women as much as I do, and many of them are on the younger side. No doubt about it --

Jeffrey enjoys his social life." Jeffrey Epstein: International Moneyman of Mystery; He's pals with a passel of Nobel Prize–winning scientists, CEOs like Leslie Wexner of the Limited, socialite Ghislaine Maxwell, even Donald Trump. But it wasn't until he flew Bill Clinton, Kevin Spacey, and Chris Tucker to Africa on his private Boeing 727 that the world began to wonder who he is. By Landon Thomas Jr. (*See* article attached hereto as Exhibit "NN") (c) Trump allegedly banned Epstein from his Maralago Club in West Palm Beach because Epstein sexually assaulted an underage girl at the club; (d) Jane Doe No. 102's complaint alleged that Jane Doe 102 was initially approached at Trump's Maralago by Ghislaine Maxwell and recruited to be Maxwell and Epstein's underage sex slave; (e) Mark Epstein (Jeffrrey Epstein's brother) testified that Trump flew on Jeffrey Epstein's plane with him (the same plane that Jane Doe 102 alleged was used to have sex with underage girls); (f) Trump had been to Epstein's home in Palm Beach; (g) Epstein's phone directory from his computer contains 14 phone numbers for Donald Trump, including emergency numbers, car numbers, and numbers to Trump's security guard and houseman. Based on this information, Edwards reasonably believed that Trump might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition. See deposition of Mark Epstein, September 21, 2009, at 48-50 (Deposition Attachment #19); *See* Jane Doe 102 v. Epstein, Exhibit "B"; Exhibit "F"; "Exhibit"J"; "N" and *See* Edwards Affidavit, Exhibit "N" at ¶13.

72. Edwards provided notice that he intended to depose Alan Dershowitz. Edwards possessed a legitimate basis for doing so: (a) Dershowitz is believed to have been friends with Epstein for many years; (b) in one news article Dershowitz comments that, "I'm on my 20th book... The only person outside of my immediate family that I send drafts to is Jeffrey" The

Talented Mr. Epstein, By Vicky Ward on January, 2005 in Published Work, Vanity Fair (*See* article attached as Exhibit "OO"); (c) Epstein's housekeeper Alfredo Rodriguez testified that Dershowitz stayed at Epstein's house during the years when Epstein was assaulting minor females on a daily basis; (d) Rodriguez testified that Dershowitz was at Epstein's house at times when underage females where there being molested by Epstein (see Alfredo Rodriguez deposition at 278-280, 385, 426-427); (e) Dershowitz reportedly assisted in attempting to persuade the Palm Beach State Attorney's Office that because the underage females alleged to have been victims of Epstein's abuse lacked credibility and could not be believed that they were at Epstein's house, when Dershowitz himself was an eyewitness to their presence at the house; (f) Jane Doe No. 102 stated generally that Epstein forced her to be sexually exploited by not only Epstein but also Epstein's "adult male peers, including royalty, politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories that Dershowitz and acquaintances of Dershowitz fall into; (g) during the years 2002-2005 Alan Dershowitz was on Epstein's plane on several occasions according to the flight logs produced by Epstein's pilot and information (described above) suggested that sexual assaults may have taken place on the plane; (h) Epstein donated $30 Million one year to the university at which Dershowitz teaches. Based on this information, Edwards had a reasonable basis to believe that Dershowitz might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition. *See* Dershowitz letters to the State Attorney's office attached as Exhibit "PP"; Deposition of Alfredo Rodriguez at 278-280; Flight Logs Exhibit "MM"; Exhibits "B" and "OO"; and Edwards Affidavit, Exhibit "N" at ¶14.

73. Epstein's complaint alleges that Edwards provided notice that he wished to take the deposition of Tommy Mattola. That assertion is untrue. Mr. Mattola's deposition was set by the law firm of Searcy Denny Scarola Barnhart and Shipley. *See* Edwards Affidavit, Exhibit "N" at ¶16.

74. Edwards gave notice that he intended to take David Copperfield's deposition. Edwards possessed a legitimate basis for doing so. Epstein's housekeeper and one of the only witnesses who did not appear for deposition with an Epstein bought attorney, Alfredo Rodriguez, testified that David Copperfield was a guest at Epstein's house on several occasions. His name also appears frequently in the message pads confiscated from Epstein's house. It has been publicly reported that Copperfield himself has had allegations of sexual misconduct made against him by women claiming he sexually abused them, and one of Epstein's sexual assault victims also alleged that Copperfield had touched her in an improper sexual way while she was at Epstein's house. Mr. Copperfield likely has relevant information and deposition was reasonably calculated to lead to the discovery of admissible evidence. *See* Edwards Affidavit, Exhibit "N" at ¶17.

75. Epstein also takes issue with Edwards identifying Bill Richardson as a possible witness. Richardson was properly identified as a possible witness because Epstein's personal pilot testified to Richardson joining Epstein at Epstein's New Mexico Ranch. There was information indicating that Epstein had young girls at his ranch which, given the circumstances of the case, raised the reasonable inference he was sexually abusing these girls as he had abused girls in West Palm Beach and elsewhere. Richardson had also returned campaign donations that were given to him by Epstein, indicating that he believed that there was something about Epstein

with which he did not want to be associated. Richardson was not called to testify nor was he ever subpoenaed to testify. *See* Edwards Affidavit, Exhibit "N" at ¶18.

76. Edwards learned of allegations that Epstein engaged in sexual abuse of minors on his private aircraft. *See* Jane Doe 102 Complaint, Exhibit "B." Accordingly, Edwards pursued discovery to confirm these allegations.

77. Discovery of the pilot and flight logs was proper in the cases brought by Edwards against Epstein. Jane Doe filed a federal RICO claim against Epstein that was an active claim through much of the litigation. The RICO claim alleged that Epstein ran an expansive criminal enterprise that involved and depended upon his plane travel. Although Judge Marra dismissed the RICO claim at some point in the federal litigation, the legal team representing Edwards' clients intended to pursue an appeal of that dismissal. Moreover, all of the subjects mentioned in the RICO claim remained relevant to other aspects of Jane Doe's claims against Epstein, including in particular her claim for punitive damages. *See* Edwards Affidavit, Exhibit "N" at ¶19.

78. Discovery of the pilot and flight logs was also proper in the cases brought by Edwards against Epstein because of the need to obtain evidence of a federal nexus. Edwards's client Jane Doe was proceeding to trial on a federal claim under 18 U.S.C. § 2255. Section 2255 is a federal statute which (unlike relevant state statutes) established a minimum level of recovery for victims of the violation of its provisions. Proceeding under the statute, however, required a "federal nexus" to the sexual assaults. Jane Doe had two grounds on which to argue that such a nexus existed to her abuse by Epstein: first, his use of telephone to arrange for girls to be abused; and, second, his travel on planes in interstate commerce. During the course of the litigation,

Edwards anticipated that Epstein would argue that Jane Doe's proof of the federal nexus was inadequate. These fears were realized when Epstein filed a summary judgment motion raising this argument. In response, the other attorneys and Edwards representing Jane Doe used the flight log evidence to respond to Epstein's summary judgment motion, explaining that the flight logs demonstrated that Epstein had traveled in interstate commerce for the purpose of facilitating his sexual assaults. Because Epstein chose to settle the case before trial, Judge Marra did not rule on the summary judgment motion.

79. Edwards had further reason to believe and did in fact believe that the pilot and flight logs might contain relevant evidence for the cases against Epstein. Jane Doe No. 102's complaint outlined Epstein's daily sexual exploitation and abuse of underage minors as young as 12 years old and alleged that Epstein's plane was used to transport underage females to be sexually abused by him and his friends. The flight logs accordingly were a potential source of information about either additional girls who were victims of Epstein's abuse or friends of Epstein who may have witnessed or even participated in the abuse. Based on this information, Edwards reasonably pursued the flight logs in discovery.

80. In the fall of 2009, Epstein gave a recorded interview to George Rush, a reporter with the *New York Daily News* about pending legal proceedings. In that interview, Epstein demonstrated an utter lack of remorse for his crimes (but indirectly admitted his crimes) by stating:

- People do not like it when people make good and that was one reason he (Epstein) was being targeted by civil suits filed by young girls in Florida;
- He (Epstein) had done nothing wrong;

- He (Epstein) had gone to jail in Florida for soliciting prostitution for no reason;
- If the same thing (i.e., sexual abuse of minor girls) had happened in New York, he (Epstein) would have received only a $200 fine;
- Bradley J. Edwards was the one causing all of Epstein's problems (i.e., the civil suits brought by Jane Doe and other girls);
- L.M. came to him as a prostitute and a drug user (i.e., came to Epstein for sex, rather than Epstein pursuing her);
- All the girls suing him are only trying to get a meal ticket;
- The only thing he might have done wrong was to maybe cross the line a little too closely;
- He (Epstein) was very upset that Edwards had subpoenaed Ghisline Maxwell, that she was a good person that did nothing wrong (i.e., had done nothing wrong even though she helped procure young girls to satisfy Epstein's sexual desires);
- With regard to Jane Doe 102 v. Epstein, which involved an allegation that Epstein had repeatedly sexually abused a 15-year-old girl, forced her to have sex with his friends, and flew her on his private plane nationally and internationally for the purposes of sexually molesting and abusing her, he (Epstein) flippantly said that the case was dismissed, indicating that the allegations were ridiculous and untrue.

See Affidavit of Michael J. Fisten attached hereto as Exhibit "QQ."

81. The Rush interview also demonstrated perjury (a federal crime) on the part of Epstein. Epstein lied about not knowing George Rush. *See* Epstein Deposition, February 17, 2010, taken in L.M. v. Jeffrey Epstein, case 50-2008-CA-028051, page 154, line 4 through 155 line 9, (Deposition attachment #7), wherein Jeffrey Epstein clearly impresses that he does not recognize George Rush from the New York Daily News. This impression was given despite the fact that he gave a lengthy personal interview about details of the case that was tape recorded with George Rush.

Epstein's Harassment of Witnesses Against Him

82. At all relevant times Edwards has a good faith basis to believe and did in fact believe that Epstein engaged in threatening witnesses. *See* Incident Report, Exhibit "A" at p. 82, U.S. Attorney's Correspondence, Exhibit "C" - Indictments drafted by Federal Government against Epstein; and Edwards Affidavit, Exhibit "N" at ¶11.

83. Despite three no contact orders entered against Epstein (*see* Exhibit C, *supra*), Edwards learned that Epstein continued to harass his victims. For example, Jane Doe had a trial set for her civil case against him on July 19, 2010. As that trial date approached, defendant Epstein intimidated her in violation of the judicial no-contact orders. On July 1, 2010, he had a "private investigator" tail Jane Doe – following her every move, stopping when she stopped, driving when she drove, refusing to pass when she pulled over. When Jane Doe ultimately drove to her home, the "private investigator" then parked in his car approximately 25 feet from Jane Doe house and flashed his high beam lights intermittently into the home. Even more threateningly, at about 10:30 p.m., when Jane Doe fled her home in the company of a retired police officer employed by Jane Doe's counsel, the "private investigator" attempted to follow Jane Doe despite a request not to do so. The retired officer successfully took evasive action and placed Jane Doe in a secure, undisclosed location that night. Other harassing actions against Jane Doe also followed. *See* Motion for Contempt filed by Edwards in Jane Doe v. Epstein detailing the event, including Fisten Affidavit attached to Motion, Composite Exhibit "RR."

Epstein Settlement of Civil Claims Against Him for Sexual Abuse of Children

84. The civil cases Edwards filed against Epstein on behalf of L.M., E.W., and Jane Doe were reasonably perceived by Edwards to be very strong cases. Because Epstein had

sexually assaulted these girls, he had committed several serious torts against them and would be liable to them for appropriate damages. *See* Preceding Undisputed Facts. Because of the outrageousness of Epstein's sexual abuse of minor girls, Edwards reasonably expected that Epstein would also be liable for punitive damages to the girls. Because Edwards could show that Epstein had molested children for years and designed a complex premeditated scheme to procure different minors everyday to satisfy his addiction to sex with minors, the punitive damages would have to be sufficient to deter him from this illegal conduct that he had engaged in daily for years. Epstein was and is a billionaire. *See* Complaint, ¶49 (referring to "Palm Beach Billionaire"); *see also* Epstein Deposition, February 17, 2010, at 172-176 (Deposition Attachment #7) (taking the Fifth when asked whether he is a billionaire). Accordingly, Edwards reasonably believed the punitive damages that would have to be awarded against Epstein would have been substantial enough to punish him severely enough for his past conduct as well as deter him from repeating his offenses in the future. *See* Edwards Affidavit, Exhibit "N" at ¶19.

85. On July 6, 2010, rather than face trial for the civil suits that had been filed against him by L.M., E.W., and Jane Doe, defendant Epstein settled the cases against him. The terms of the settlement are confidential. The settlement amounts are highly probative in the instant action as Epstein bases his claims that Edwards was involved in the Ponzi scheme on Epstein's inability to settle the L.M., E.W., and Jane Doe cases for "minimal value". His continued inability to settle the claims for "minimal value" after the Ponzi scheme was uncovered would be highly probative in discrediting any causal relationship between the Ponzi scheme and Edwards's settlement negotiations. *See* Edwards Affidavit, Exhibit "N" at ¶21.

Edwards Non-Involvement in Fraud by Scott Rothstein

86. From in or about 2005, through in or about November 2009, Scott Rothstein appears to have run a giant Ponzi scheme at his law firm of Rothstein, Rosenfeldt and Adler P.A. ("RRA"). This Ponzi scheme involved Rothstein falsely informing investors that settlement agreements had been reached with putative defendants based upon claims of sexual harassment and/or whistle-blower actions. Rothstein falsely informed the investors that the potential settlement agreements were available for purchase. Plea Agreement at 2, *United States v. Scott W. Rothstein*, No. 9-60331-CR-COHN (S.D. Fla. Jan. 27, 2010) attached hereto as Exhibit "SS."

87. It has been alleged that among other cases that Rothstein used to lure investors into his Ponzi scheme were the cases against Epstein that were being handled by Bradley J. Edwards, Esq. Edwards had no knowledge of the fraud or any such use of the Epstein cases. *See* Edwards Affidavit, Exhibit "N" at ¶9.

88. Bradley J. Edwards, Esq., joined RRA in about April 2009 and left RRA in November 2009 – a period of less than one year. Edwards would not have joined RRA had he been aware that Scott Rothstein was running a giant Ponzi scheme at the firm. Edwards left RRA shortly after learning of Rothstein's fraudulent scheme. *Id.* at ¶8.

89. At no time prior to the public disclosure of Rothstein's Ponzi scheme did Edwards know or have reason to believe that Rothstein was using legitimate claims that Edwards was prosecuting against Epstein for any fraudulent or otherwise illegitimate purpose. *Id.* at ¶20.

90. Edwards never substantively discussed the merits of any of his three cases against Epstein with Rothstein. *See* Deposition of Bradley J. Edwards taken March 23, 2010, at 110-16. (hereinafter "Edwards Depo") (Deposition Attachment #22).

91. On July 20, 2010, Bradley Edwards received a letter from the U.S. Attorney's Office for the Southern District of Florida – the office responsible for prosecuting Rothstein's Ponzi scheme. The letter indicated that law enforcement agencies had determined that Edwards was "a victim (or potential victim)" of Scott Rothstein's federal crimes. The letter informed Edwards of his rights as a victim of Rothstein's fraud and promised to keep Edwards informed about subsequent developments in Rothstein's prosecution. *See* Letter attached hereto as Exhibit "TT."

92. Jeffrey Epstein filed a complaint with the Florida Bar against Bradley Edwards, Esq., raising allegations that Edwards and others were involved in the wrongdoing of Scott Rothstein. After investigating the claim, the Florida Bar dismissed this complaint. *See* Edwards Affidavit, Exhibit "N" at ¶23.

Epstein Takes the Fifth When Asked Substantive Questions About His Claims Against Edwards

93. On March 17, 2010, defendant Epstein was deposed about his lawsuit against Edwards. Rather than answer substantive questions about his lawsuit, Epstein repeatedly invoked his Fifth Amendment privilege. *See* Epstein Depo. taken 3/17/10, Deposition Attachment #1.

94. In his deposition, Epstein took the Fifth rather than answer the question: "Specifically what are the allegations against you which you contend Mr. Edwards ginned up?" *Id.* at 34.

95. In his deposition, Epstein took the Fifth rather than name people in California that Edwards had tried to depose to increase the settlement value of the civil suit he was handling. *Id.* at 37.

96. In his deposition, Epstein took the Fifth rather than answer the question: "Do you know former President Clinton personally." *Id.*

97. In his deposition, Epstein took the Fifth rather than answer the question: "Are you now telling us that there were claims against you that were fabricated by Mr. Edwards?" *Id.* at 39.

98. In his deposition, Epstein took the Fifth rather than answer the question, "Well, which of Mr. Edwards' cases do you contend were fabricated." *Id.*

99. In his deposition, Epstein took the Fifth rather than answer the question: "What is the actual value that you contend the claim of E.W. against you has?" *Id.* at 45.

100. In his deposition, Epstein took the Fifth rather than answer a question about the actual value of the claim of L.M. and Jane Doe against him. *Id.*

101. In his deposition, taken prior to the settlement of Edwards's clients claims against Epstein, Epstein took the Fifth rather than answer the question: "Is there any pending claim against you which you contend is fabricated?" *Id.* at 71.

102. In his deposition, Epstein took the Fifth rather than answer the question: "Did you ever have damaging evidence in your garbage?" *Id.* at 74.

103. In his deposition, Epstein took the Fifth rather than answer the question: "Did sexual assaults ever take place on a private airplane on which you were a passenger?" *Id.* at 88.

104. In his deposition, Epstein took the Fifth rather than answer the question: "Does a flight log kept for a private jet used by you contain the names of celebrities, dignitaries or international figures?" *Id.* at 89.

105. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever socialized with Donald Trump in the presence of females under the age of 18?" *Id.* at 89.

106. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever socialized with Alan Dershowitz in the presence of females under the age of 18." *Id.* at 90.

107. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever socialized with Mr. Mottola in the presence of females under the age of 18?" *Id.* at 91-92.

108. In his deposition, Epstein took the Fifth rather than answer the question: "Did you ever socialize with David Copperfield in the presence of females under the age of 18?" *Id.* at

109. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever socialized with Mr. Richardson [Governor of New Mexico and formerly U.S. Representative and Ambassador to the United Nations] in the presence of females under the age of 18." *Id.* at 94.

110. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever sexually abused children?" *Id.* at 95.

111. In his deposition, Epstein took the Fifth rather than answer the question: "Did you have staff members that assisted you in scheduling appointments with underage females; that is, females under the age of 18." *Id.* at 97-98.

112. In his deposition, Epstein took the Fifth rather than answer the question: "On how many occasions did you solicit prostitution." *Id.* at 102.

113. In his deposition, Epstein took the Fifth rather than answer the question: "How many minors have you procured for prostitution?" *Id.* at 104.

114. In his deposition, Epstein took the Fifth rather than answer the question: "Have you ever coerced, induced or enticed any minor to engage in any sexual act with you?" *Id.* at 107.

115. In his deposition, Epstein took the Fifth rather than answer the question: "How many times have you engaged in fondling underage females?" *Id.* at 108.

116. In his deposition, Epstein took the Fifth rather than answer the question: "How many times have you engaged in oral sex with females under the age of 18?" *Id.* at 110.

117. In his deposition, Epstein took the Fifth rather than answer the question: "Do you have a personal sexual preference for children?" *Id.* at 111-12.

118. In his deposition, Epstein took the Fifth rather than answer the question: "Your Complaint at page 27, paragraph 49, says that 'RRA and the litigation team took an emotionally driven set of facts involving alleged innocent, unsuspecting, underage females and a Palm Beach billionaire, and sought to turn it into a goldmine,' end of quote. Who is the Palm Beach billionaire referred to in that sentence?" *Id.* at 112-13.

119. In his deposition, Epstein took the Fifth rather than answer the question: "Who are the people who are authorized to make payment [to your lawyers] on your behalf?" *Id.* at 120.

120. In his deposition, Epstein took the Fifth rather than answer the question: "Is there anything in L.M.'s Complaint that was filed against you in September of 2008 which you contend to be false?" *Id.* at 128.

CERTIFICATE OF SERVICE

I HEREBY CERTIFY that on November 4th, 2010 a copy of the foregoing has been served via Fax and U.S. Mail to all those on the attached service list.

Jack Scarola
Searcy, Denney, Scarola, Barnhart & Shipley
2139 Palm Beach Lakes Blvd
West Palm Beach, FL 33409
(561) 686-6300
(561) 684-5816 (fax)

By: ______________________________
JACK SCAROLA
Florida Bar No.: 169440

SERVICE LIST

Christopher E. Knight, Esq.
Joseph L. Ackerman, Esq.
FOWLER WHITE BURNETT P.A.
901 Phillips Point West
777 South Flagler Drive
West Palm Beach, FL 33401

Jack Alan Goldberger, Esq.
Atterbury Goldberger et al.
250 Australian Avenue South
Suite 1400
West Palm Beach, FL 33401

Marc S. Nurik, Esq.
Law Offices of Marc S. Nurik
One E. Broward Blvd., Suite 700
Fort Lauderdale, FL 33301

Gary M. Farmer, Jr.
Farmer, Jaffe, Weissing,
Edwards, Fistos & Lehrman, P.L.
425 N. Andrews Ave., Suite 2
Fort Lauderdale, FL 33301

EXHIBIT N

AFFIDAVIT OF BRADLEY JAMES EDWARDS

1. I am an attorney in good standing with the Florida Bar and admitted to practice in the Southern District of Florida. I am currently a partner in the law firm of Farmer, Jaffe, Weissing, Edwards, Fistos & Lehrman, P.L.

2. In 2008, I was a sole practitioner running a personal injury law firm in Hollywood, FL. While a sole practitioner I was retained by three clients, L.M., E.W., and Jane Doe to pursue civil litigation against Jeffrey Epstein for sexually abusing them while they were minor girls. I agreed to represent these girls, along with attorney Jay Howell (an attorney in Jacksonville, Florida with Jay Howell & Associates) and Professor Paul Cassell (a law professor at the University of Utah College Of Law). I filed state court actions on behalf of L.M. and E.W. and a federal court action on behalf of Jane Doe. All of the cases were filed in the summer of 2008.

3. My clients received correspondence from the U.S. Department of Justice regarding their rights as victims of Epstein's federal sex offenses. (True and accurate copies of the letters are attached to Statement of Undisputed Facts as Exhibit "M")

4. In mid June 2008, I contacted Assistant United States Attorney Marie Villafaña to inform her that I represented Jane Doe #1(E.W.) and, later, Jane Doe #2(L.M.). I asked to meet to provide information regarding Epstein. AUSA Villafaña did not advise me that a plea agreement had already been negotiated with Epstein's attorneys that would block federal prosecution. AUSA Villafaña did indicate that federal investigators had concrete evidence and information that Epstein had sexually molested at least 40 underage minor females, including E.W., Jane Doe and L.M.

5. I also requested from the U.S. Attorney's Office the information and evidence that they had collected regarding Epstein's sexual abuse of his clients. However, the U.S. Attorney's Office declined to provide any such information to me. The U.S. Attorney's Office also declined to provide any such information to the other attorneys who represented victims of Epstein's sexual assaults.

6. I was informed that on Friday, June 27, 2008, at approximately 4:15 p.m., AUSA Villafaña received a copy of Epstein's proposed state plea agreement and learned that the plea was scheduled for 8:30 a.m., Monday, June 30, 2008. She called me to provide notice to my clients regarding the hearing. She did not tell me that the guilty pleas in state court would bring an end to the possibility of federal prosecution pursuant to the plea agreement. My clients did not learn and understand this fact until July 11, 2008, when the agreement was described during a hearing held before Judge Marra on the Crime Victims' Rights Act action that I had filed.

7. In the summer of 2008 I filed complaints against Jeffrey Epstein on behalf of L.M., E.W., and Jane Doe.

8. In the Spring of 2009 (approximately April), I joined the law firm of Rothstein, Rosenfeldt and Adler, P.A. ("RRA"). I brought my existing clients with me when I joined RRA, including L.M., E.W., and Jane Doe. When I joined the firm, I was not aware that Scott Rothstein was running a Ponzi scheme at RRA. Had I known such a Ponzi scheme was in place, I would never have joined RRA.

9. I am now aware that it has been alleged that Scott Rothstein made fraudulent presentations to investors about the lawsuits that I had filed on behalf of my clients against Epstein and that it has been alleged that these lawsuits were used to fraudulently lure investors into Rothstein's Ponzi scheme. I never met a single investor, had no part in any such presentations and had no knowledge any such fraud was occurring. If these allegations are true, I had no knowledge that any such fraudulent presentations were occurring and no knowledge of any such improper use of the case files.

10. Epstein's Complaint against me alleges that Rothstein made false statements about cases filed against Epstein, i.e., that RRA had 50 anonymous females who had filed suit against Epstein; that Rothstein sold an interest in personal injury lawsuits, reached agreements to share attorneys fees with non-lawyers, paid clients "up front" money; and that he used the judicial process to further his Ponzi scheme. If Rothstein did any of these things, I had no knowledge of his actions. Because I maintained close contact with my clients, EW, LM and Jane Doe, and Scott Rothstein never met any of them, I know for certain that none of my clients were paid "up front" money by anyone.

11. Epstein alleges that I attempted to take the depositions of his "high profile friends and acquaintances" for no legitimate litigation purpose. This is untrue, as all of my actions in representing L.M., E.W., and Jane Doe were aimed at providing them effective representation in their civil suits. With regard to Epstein's friends, through documents and information obtained in discovery and other means of investigation, I learned that Epstein was sexually molesting minor girls on a daily basis and had been for many years. I also learned the unsurprising fact that he was molesting the girls in the privacy of his mansion in West Palm Beach, meaning that locating witnesses to corroborate their testimony would be difficult to find. I also learned, from the course of the litigation, that Epstein and his lawyers were constantly attacking the credibility of the girls, that Epstein's employees were all represented by lawyers who apparently were paid for (directly or indirectly) by Epstein, that co-conspirators whose representation was also apparently paid for by Epstein were all taking the Fifth (like Epstein) rather than provide information in discovery. For example, I was given reason to believe that Sarah Kellen, Larry Visoski, Larry Harrison, David Rogers, Louella Rabuyo, Nadia Marcinkova, Ghislaine Maxwell, Mark Epstein, and Janusz Banasiak all had lawyers paid for by Epstein. Because Epstein and the co-conspirators in his child molestation criminal enterprise blocked normal discovery avenues, I needed to search for other ordinary approaches to strengthen the cases of my clients. Consistent with my training and experience, these other ordinary approaches included finding other witnesses who could corroborate allegations of sexual abuse of my clients or other girls. Some of these witnesses were friends of Epstein. Given his social status, it also turned out that some of his friends were high-profile individuals.

12. In light of information I received suggesting that British socialite Ghislaine Maxwell, former girlfriend and long-time friend of Epstein's, was involved in managing Epstein's affairs and companies I had her served for deposition for August 17, 2009. (Deposition Notice attached to Statement of Undisputed Facts as Exhibit BB). Maxwell was represented by Brett Jaffe of the New York firm of Cohen and Gresser, and I understood that her attorney was paid for (directly or indirectly) by Epstein. She was reluctant to give her deposition, and I tried to work with her attorney to take her deposition on terms that would be acceptable to both sides. Her attorney and I negotiated a confidentiality agreement, under which Maxwell agreed to drop any objections to the deposition. Maxwell, however, still avoided the deposition. On June 29, 2010, one day before I was to fly to NY to take Maxwell's deposition, her attorney informed me that Maxwell's mother was deathly ill and Maxwell was consequently flying to England with no intention of returning and certainly would not return to the United States before the conclusion of Jane Doe's trial period (August 6, 2010). Despite that assertion, I later learned that Ghislaine Maxwell was in fact in the country on approximately July 31, 2010, as she attended the wedding of Chelsea Clinton (former President Clinton's daughter) and was captured in a photograph taken for US Weekly magazine.

13. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Donald Trump's deposition in the civil suits against him. Trump was properly noticed because: (a) after review of the message pads confiscated from Epstein's home, the legal and investigative team assisting my clients learned that Trump called Epstein's West Palm Beach mansion on several occasions during the time period most relevant to my clients' complaints; (b) Trump was quoted in a *Vanity Fair* article about Epstein as saying "**I've known Jeff for fifteen years. Terrific guy.**" "**He's a lot of fun to be with. It is even said that he likes beautiful women as much as I do, and many of them are on the younger side. No doubt about it -- Jeffrey enjoys his social life.**" Jeffrey Epstein: International Moneyman of Mystery; He's pals with a passel of **Nobel Prize–winning scientists, CEOs like Leslie Wexner of the Limited, socialite Ghislaine Maxwell, even Donald Trump. But it wasn't until he flew Bill Clinton, Kevin Spacey, and Chris Tucker to Africa on his private Boeing 727 that the world began to wonder who he is. By Landon Thomas Jr.**; (c) I learned through a source that Trump banned Epstein from his Maralago Club in West Palm Beach because Epstein sexually assaulted an underage girl at the club; (d) Jane Doe No. 102's complaint alleged that Jane Doe 102 was initially approached at Trump's Maralago by Ghislaine Maxwell and recruited to be Maxwell and Epstein's underage sex slave; (e) Mark Epstein (Jeffrey Epstein's brother) testified that Trump flew on Jeffrey Epstein's plane with him (the same plane that Jane Doe 102 alleged was used to have sex with underage girls) deposition of Mark Epstein, September 21, 2009 at 48-50; (f) Trump visited Epstein at his home in Palm Beach – the same home where Epstein abused minor girls daily; (g) Epstein's phone directory from his computer contains 14 phone numbers for Donald Trump, including emergency numbers, car numbers, and numbers to Trump's security guard and houseman. Based on this information, I believed that

Trump might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

14. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Alan Dershowitz's deposition in the civil suits against him. Dershowitz was properly noticed because: (a) Dershowitz has been friends with Epstein for many years; (b) in one news article Dershowitz comments that, "I'm on my 20th book... The only person outside of my immediate family that I send drafts to is Jeffrey" The Talented Mr. Epstein, By Vicky Ward on January, 2005 in Published Work, Vanity Fair; (c) Epstein's housekeeper Alfredo Rodriguez testified that Dershowitz stayed at Epstein's house during the years most relevant to my clients; (d) Rodriguez testified that Dershowitz was at Epstein's house at times when underage females where there being molested by Epstein (see Alfredo Rodriguez deposition at 278-280, 385, 426-427); (e) Dershowitz was reportedly involved in persuading the Palm Beach State Attorney's office not to file felony criminal charges against Epstein because the underage females lacked credibility and thus could not be believed that they were at Epstein's house, despite him being an eyewitness that the underage girls were actually there; (f) Jane Doe No. 102 stated generally that Epstein forced her to be sexually exploited by not only Epstein but also Epstein's "adult male peers, including royalty, politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories that Dershowitz and acquaintances of Dershowitz fall into; (g) during the years 2002-2005 Alan Dershowitz was on Epstein's plane on several occasions according to the flight logs produced by Epstein's pilot and information (described above) suggested that sexual assaults may have taken place on the plane; (h) Epstein donated Harvard $30 Million dollars one year, and Harvard was one of the only institutions that did not return Epstein's donation after he was charged with sex offenses against children. Based on this information, I believed that Dershowitz might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

15. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Bill Clinton's deposition. Clinton was properly noticed because: (a) it was well known that Clinton was friends with Ghislaine Maxwell, and several witnesses had provided information that Maxwell helped to run Epstein's companies, kept images of naked underage children on her computer, helped to recruit underage children for Epstein, engaged in lesbian sex with underage females that she procured for Epstein, and photographed underage females in sexually explicit poses and kept child pornography on her computer; (b) newpaper articles stated that Clinton had an affair with Ghislaine Maxwell, who was thought to be second in charge of Epstein's child molestation ring. The Cleveland Leader newspaper, April 10, 2009; (c) it was national news when Clinton traveled with Epstein (and Maxwell) aboard Epstein's private plane to Africa and the news articles classified Clinton as Epstein's friend; (d) the flight logs for the relevant years 2002 - 2005 showed Clinton traveling on Epstein's plane on more than 10 occasions and his assistant, Doug Band, traveled on many more occasions; (e) Jane Doe No. 102 stated generally that she was required by Epstein to be sexually

exploited by not only Epstein but also Epstein's "adult male peers, including royalty, politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories Clinton and acquaintances of Clinton fall into; (f) flight logs showed that Clinton took many flights with Epstein, Ghislaine Maxwell, Sarah Kellen, and Adriana Mucinska, -- all employees and/or co-conspirators of Epstein's that were closely connected to Epstein's child exploitation and sexual abuse; (g) Clinton frequently flew with Epstein aboard his plane, then suddenly stopped – raising the suspicion that the friendship abruptly ended, perhaps because of events related to Epstein's sexual abuse of children; (h) Epstein's personal phone directory from his computer contains e-mail addresses for Clinton along with 21 phone numbers for him, including those for his assistant (Doug Band), his schedulers, and what appear to be Clinton's personal numbers. Based on this information, I believed that Clinton might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

16. Epstein alleges that Tommy Mottola was improperly noticed with a deposition. I did not notice Mattola for deposition. He was noticed for deposition by a law firm representing another one of Epstein's victims – not by me.

17. Epstein alleges that there was something improper in the fact that I notified him that I intended to take the illusionist David Copperfield's deposition. Copperfield was properly noticed because: (a) Epstein's housekeeper Alfredo Rodriguez testified that David Copperfield was a guest on several occasions at Epstein's house; (b) according to the message pads confiscated from Epstein's house, Copperfield called Epstein quite frequently and left messages that indicated they socialized together; (c) Copperfield himself has had similar allegations made against him by women claiming he sexually abused them; (d) one of Epstein's sexual assault victims also alleged that Copperfield had touched her in an improper sexual way while she was at Epstein's house. Based on this information, I believed that Copperfield might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

18. Epstein alleges that there was something improper in the fact that I identified Bill Richardson as a possible witness against him in the civil cases. Richardson was properly identified as a possible witness because Epstein's personal pilot testified to Richardson joining Epstein at Epstein's New Mexico Ranch. See deposition of Larry Morrison, October 6, 2009, at 167-169. There was information indicating that Epstein had young girls at his ranch which, given the circumstances of the case, raised the reasonable inference he was sexually abusing these girls since he had regularly and frequently abused girls in West Palm Beach and elsewhere. Richardson had also returned campaign donations that were given to him by Epstein, indicating that he believed that there was something about Epstein that he did not want to be associated with. Richardson was not called to testify nor was he ever subpoenaed to testify.

19. Epstein alleges that discovery of plane and pilot logs was improper during discovery in the civil cases against him. Discovery of these subjects was clearly proper and

necessary because: (a) Jane Doe filed a federal RICO claim against Epstein that was an active claim through much of the litigation. The RICO claim alleged that Epstein ran an expansive criminal enterprise that involved and depended upon his plane travel. Although Judge Marra dismissed the RICO claim at some point in the federal litigation, the legal team representing my clients intended to pursue an appeal of that dismissal. Moreover, all of the subjects mentioned in the RICO claim remained relevant to other aspects of Jane Doe's claims against Epstein, including in particular her claim for punitive damages; (b) Jane Doe also filed and was proceeding to trial on a federal claim under 18 U.S.C. § 2255. Section 2255 is a federal statute which (unlike other state statutes) guaranteed a minimum level of recovery for Jane Doe. Proceeding under the statute, however, required a "federal nexus" to the sexual assaults. Jane Doe had two grounds on which to argue that such a nexus existed to her abuse by Epstein: first, his use of the telephone to arrange for girls to be abused; and, second, his travel on planes in interstate commerce. During the course of the litigation, I anticipated that Epstein would argue that Jane Doe's proof of the federal nexus was inadequate. These fears were realized when Epstein filed a summary judgment motion raising this argument. In respo-nse, the other attorneys and I representing Jane Doe used the flight log evidence to respond to Epstein's summary judgment motion, explaining that the flight logs demonstrated that Epstein had traveled in interstate commerce for the purpose of facilitating his sexual assaults. Because Epstein chose to settle the case before trial, Judge Marra did not rule on the summary judgment motion. (c) Jane Doe No. 102's complaint outlined Epstein's daily sexual exploitation and abuse of underage minors as young as 12 years old and alleged that he used his plane to transport underage females to be sexually abused by him and his friends. The flight logs accordingly might have information about either additional girls who were victims of Epstein's abuse or friends of Epstein who may have witnessed or even participated in the abuse. Based on this information, I believed that the flight logs and related information was relevant information to prove the cases against Jeffrey Epstein and accordingly I pursued them in discovery.

20. In approximately November 2009, the existence of Scott Rothstein's Ponzi scheme became public knowledge. It was at that time that I, along with many other reputable attorneys at RRA, first became aware of Rothstein criminal scheme. At that time, I left RRA with several other RRA attorneys to form the law firm of Farmer Jaffe Weissing Edwards Fistos and Lehrman ("Farmer Jaffe"). I was thus with RRA for less than one year.

21. In July 2010, along with other attorneys at Farmer Jaffe and Professor Cassell, I reached favorable settlement terms for my three clients L.M., E.W., and Jane Doe in their lawsuits against Epstein.

22. On July 20, 2010, I received a letter from the U.S. Attorney's Office for the Southern District of Florida – the office responsible for prosecuting Rothstein's Ponzi scheme. The letter indicated that law enforcement agencies had determined that I was "a victim (or potential victim)" of Scott Rothstein's federal crimes. The letter informed me of my rights as a victim of Rothstein's federal crimes and promised to keep me informed about

subsequent developments in his prosecution. A copy of this letter is attached to this Affidavit. (A copy of the letter is attached to Statement of Undisputed Facts as Exhibit UU)

23. Jeffrey Epstein also filed a complaint with the Florida Bar against me. His complaint alleged that I had been involved in Rothstein's scheme and had thereby violated various rules of professional responsibility. The Florida Bar investigated and dismissed the complaint.

24. I have reviewed the Statement of Undisputed Facts filed contemporaneously with this Affidavit. Each of the assertions concerning what I learned, what I did, and the good faith beliefs formed by me in the course of my prosecutions of claims against Jeffrey Epstein as contained in the Statement of Undisputed Facts is true, and the foundations set out as support for my beliefs are true and correct to the best of my knowledge.

25. All actions taken by me in the course of my prosecution of claims against Jeffrey Epstein were based upon a good faith belief that they were reasonable, necessary, and ethically proper to fulfill my obligation to zealously represent the interests of my clients.

I declare under penalty of perjury that the foregoing is true and correct.

Dated: 9/21, 2010

Bradley J. Edwards, Esq.

IN THE CIRCUIT COURT OF THE 15TH
JUDICIAL CIRCUIT IN AND FOR PALM
BEACH COUNTY, FLORIDA

Case No.: 50 2009CA 040800XXXXMBAG

JEFFREY EPSTEIN,

Plaintiff,

vs.

SCOTT ROTHSTEIN, individually, and
BRADLEY J. EDWARDS, individually,

Defendants,
_______________________________/

DEFENDANT BRADLEY J. EDWARDS'S RENEWED MOTION FOR FINAL SUMMARY JUDGMENT

Defendant, Bradley J. Edwards, Esq., by and through his undersigned counsel and pursuant to Rule 1.510, Florida Rules of Civil Procedure, hereby moves for Final Summary Judgment and in support thereof states as follows:

I. INTRODUCTION

The pleadings and discovery taken to date show that there is no genuine issue as to any material facts and that Bradley J. Edwards, Esq. is entitled to summary judgment for all claims brought against him in Plaintiff Jeffrey Epstein's Second Amended Complaint. Not only is there an absence of competent evidence to demonstrate that Edwards participated in any fraud against Epstein, the evidence uncontrovertibly demonstrates the propriety of every aspect of Edwards' involvement in the prosecution of legitimate claims against Epstein. Epstein sexually abused three clients of Edwards – L.M., E.W., and Jane Doe – and Edwards properly and successfully

EXHIBIT
B

represented them in a civil action against Epstein. Nothing in Edwards's capable and competent representation of his clients can serve as the basis for a civil lawsuit against him. Allegations about Edwards's participation in or knowledge of the use of the civil actions against Epstein in a "Ponzi Scheme" are not supported by any competent evidence and could never be supported by competent evidence as they are entirely false.

A. Epstein's Complaint

Epstein's Second Amended Complaint essentially alleges that Epstein was damaged by Edwards, acting in concert with Scott Rothstein (President of the Rothstein Rosenfeldt Adler law firm ("RRA") where Edwards worked for a short period of time). Epstein appears to allege that Edwards joined Rothstein in the abusive prosecution of sexual assault cases against Epstein to "pump" the cases to Ponzi scheme investors. As described by Epstein, investor victims were told by Rothstein that three minor girls who were sexually assaulted by Epstein: L.M., E.W., and Jane Doe were to be paid up-front money to prevent those girls from settling their civil cases against Epstein. In Epstein's view, these child sexual assault cases had "minimal value" (Complaint & 42(h)), and Edwards's refusal to force his clients to accept modest settlement offers is claimed to breach some duty that Edwards owed to Epstein. Interestingly, Epstein never states that he actually made any settlement offers.

The supposed "proof" of the Complaint's allegations against Edwards includes Edwards's alleged contacts with the media, his attempts to obtain discovery from high-profile persons with whom Epstein socialized, and use of "ridiculously inflammatory" language in arguments in court. Remarkably, Epstein has filed such allegations against Edwards despite the fact that Epstein had sexually abused each of Edwards's clients and others while they were minors. Indeed, in discovery Epstein has asserted his Fifth Amendment privilege rather than answer questions about the extent of the sexual abuse of his many victims. Even more remarkably, since filing his suit against Edwards, Epstein has now settled the three cases Edwards handled for an amount that Epstein insisted be kept confidential. Without violating the strict confidentiality terms required by Epstein, the cases did not settle for the "minimal value" that Epstein suggested in his Complaint. Because Epstein relies upon the alleged discrepancy between the "minimal value" Epstein ascribed to the claims and the substantial value Edwards sought to recover for his clients, the settlement amounts Epstein voluntarily agreed to pay while these claims against Edwards were pending will be disclosed to the court in-camera.

B. Summary of the Argument

Bradley J. Edwards, Esq., is entitled to summary judgment on Epstein's frivolous claim for at least three separate reasons.

First, because Epstein has elected to hide behind the shield of his right against self incrimination to preclude his disclosing any relevant information about the criminal activity at the center of his claims, he is barred from prosecuting this case against Edwards. Under the well-established "sword and shield" doctrine, Epstein cannot seek damages from Edwards while at the same time asserting a Fifth Amendment privilege to block relevant discovery. His case must therefore be dismissed.

Second, all of Edwards' conduct in the prosecution of valid claims against Epstein is protected by the litigation privilege.

Third, and most fundamentally, Epstein's lawsuit should be dismissed because it is not only unsupported by but is also directly contradicted by all of the record evidence. From the beginning, Edwards diligently represented three victims of sexual assaults perpetrated by Epstein. As explained in detail below, each and every one of Edwards's litigation decisions was grounded in proper litigation judgment about the need to pursue effective discovery against Epstein, particularly in the face of Epstein's stonewalling tactics. Edwards's successful representation finally forced Epstein to settle and pay appropriate damages. Effective and proper representation of child victims who have been repeatedly sexually assaulted cannot form the basis of a separate, "satellite" lawsuit, and therefore Edwards is entitled to summary judgment on these grounds as well.

The truth is the record is entirely devoid of any evidence to support Epstein's claims and is completely and consistently corroborative of Edwards's sworn assertion of innocence. Put simply, Epstein has made allegations that have no basis in fact. To the contrary, his lawsuit was merely a desperate measure by a serial pedophile to prevent being held accountable for repeatedly sexually abusing minor females. Epstein's ulterior motives in filing and prosecuting this lawsuit are blatantly obvious. Epstein's behavior is another clear demonstration that he feels he lives above the law and that because of his wealth he can manipulate the system and pay for lawyers to do his dirty work - even to the extent of having them assert baseless claims against other members of the Florida Bar. Epstein's Second Amended Complaint against Edwards is nothing short of a far-fetched fictional fairy-tale with absolutely no evidence whatsoever to support his preposterous claims. It was his last ditch effort to escape the public disclosure by Edwards and his clients of the nature, extent, and sordid details of his life as a serial child molester. Edwards's Motion for Summary Judgment should be granted without equivocation.

ARGUMENT

II. EDWARDS IS ENTITLED TO SUMMARY JUDGMENT ON EPSTEIN'S CLAIM BECAUSE THERE ARE NO MATERIAL DISPUTED FACTS AND THE UNDISPUTED FACTS ESTABLISH THAT EDWARDS'S CONDUCT COULD NOT POSSIBLY FORM THE BASIS OF ANY LIABILITY IN FAVOR OF EPSTEIN

A. The Summary Judgment Standard.

Rule 1.510(c), Florida Rules of Civil Procedure, provides that a court may enter summary judgment when the pleadings, depositions and factual showings reveal that there is no genuine issue of material fact and that the moving party is entitled to judgment as a matter of law. *See Snyder v. Cheezem Development Corp.*, 373 So. 2d 719, 720 (Fla. 2d DCA 1979); Rule 1.510(c), Fla. R. Civ. P. Once the moving party conclusively establishes that the nonmoving party cannot prevail, it is incumbent on the nonmoving party to submit evidence to rebut the motion for summary judgment. *See Holl v. Talcott,* 191 So. 2d 40, 43 (Fla. 1966). It is not enough for the opposing party merely to assert that an issue of fact does exist. *Fisel v. Wynns*, 667 So.2d 761, 764 (Fla.1996); *Landers v. Milton,* 370 So.2d 368, 370 (Fla.1979) (same).

Moreover, it is well-recognized that the non-moving party faced with a summary judgment motion supported by appropriate proof may not rely on bare, conclusory assertions found in the pleadings to create an issue and thus avoid summary judgment. Instead, the party must produce counter-evidence establishing a genuine issue of material fact. *See Bryant v. Shands Teaching Hospital and Clinics, Inc.,* 479 So.2d 165, 168 (Fla. 1st Dist. Ct. App. 1985); *see also Lanzner v. City of North Miami Beach,* 141 So.2d 626 (Fla. 3d Dist Ct. App. 1962) (recognizing that mere contrary allegations of complaint were not sufficient to preclude summary judgment on basis of facts established without dispute). Where the nonmoving party fails to present evidence rebutting the motion for summary judgment and there is no genuine issue of material fact, then entry of judgment is proper as a matter of law. *See Davis v. Hathaway*, 408 So. 2d 688, 689 (Fla. 2d Dist. Ct. App. 1982); *see also Holl*, 191 So. 2d at 43.

B. Epstein's Claim Regarding Edwards Have Absolutely No Factual Basis.

This is not a complicated case for granting summary judgment. To the contrary, this is a simple case for summary judgment because each and every one of Epstein's claim against Edwards lacks any merit whatsoever.[1]

1. Epstein's allegations regarding Edwards' involvement in Rothstein's "Ponzi Scheme" are unsupported and unsupportable because he was simply not involved in any such scheme.

a. Edwards Had No Involvement in the Ponzi Scheme.

The bulk of Epstein's claims against Edwards hinge on the premise that Edwards was involved in a Ponzi scheme run by Scott Rothstein. Broad allegations of wrongdoing on the part of Edwards are scattered willy-nilly throughout the complaint. None of the allegations provide any substance as to how Edwards actually assisted the Ponzi scheme, and allegations that he "knew or should have known" of its existence are based upon an impermissible pyramiding of inferences. In any event, these allegations all fail for one straightforward reason: Edwards was simply not involved in any Ponzi scheme. He has provided sworn testimony and an affidavit in support of that assertion, and there is not (and could never be) any contrary evidence.

Edwards has now been deposed at length in this case. As his deposition makes crystal clear, he had no knowledge of any fraudulent activity in which Scott Rothstein may have been involved. *See, e.g.*, Edwards Depo. at 301-02 (Q: " . . . [W]ere you aware that Scott Rothstein was trying to market Epstein cases . . . ?" A: "No.").

---

[1] A decision by the Court to grant summary judgment on Epstein's claims against Edwards would not affect Epstein's claims against Scott Rothstein. Epstein has already chosen to dismiss all of his claims against L.M., the only other defendant named in the suit.

Edwards has supplemented his deposition answers with an Affidavit that declares in no uncertain terms his lack of involvement in any fraud perpetrated by Rothstein. *See, e.g.*, Edwards Affidavit attached to Statement of Undisputed Material Facts as Exhibit "N" at ¶8-10, ¶20, ¶22-23. Indeed, no reasonable juror could find that Edwards was involved in the scheme, as Edwards joined RRA well after Rothstein began his fraud and would have been already deeply in debt. In fact, the evidence of Epstein's crimes is now clear, and Edwards's actions in this case were entirely in keeping with his obligation to provide the highest possible quality of legal representation for his clients to obtain the best result possible.

In view of this clear evidence rebutting all allegations against him, Epstein must now "produce counter-evidence establishing a genuine issue of material fact." *See Bryant v. Shands Teaching Hospital and Clinics, Inc.*, 479 So.2d 165, 168 (Fla. 1st Dist. Ct. App. 1985). Epstein cannot do this. Indeed, when asked at his deposition whether he had any evidence of Edwards's involvement, Epstein declined to answer, purportedly on attorney-client privilege grounds:

> Q. I want to know whether you have any knowledge of evidence that Bradley Edwards personally ever participated in devising a plan through which were sold purported confidential assignments of a structured payout settlement? . . .
> A. I'd like to answer that question by saying that the newspapers have reported that his firm was engaged in fraudulent structured settlements in order to fleece unsuspecting Florida investors. With respect to my personal knowledge, I'm unfortunately going to, today, but I look forward to at some point being able to disclose it, today I'm going to have to assert the attorney/client privilege.

See Deposition of Jeffrey Epstein, Mar. 17, 2010 (hereinafter "Epstein Depo.") at 67-68. Therefore summary judgment should be granted for Edwards on all claims involving any Ponzi scheme by Rothstein.

b. Epstein's Allegations of Negligence by Edwards are Unfounded and Not Actionable in Any Event.

In his Second Amended Complaint Epstein recognizes at least the possibility that Edwards was not involved in any Rothstein Ponzi scheme. Therefore, seemingly as a fallback, Epstein alleges without explanation that Edwards "should have known" about the existence of this concealed Ponzi scheme. Among other problems, this fallback negligence position suffers the fatal flaw that it does not link at all to the intentional tort of abuse of process alleged in the complaint.

Epstein's negligence claim is also deficient because it simply fails to satisfy the requirements for a negligence cause of action:

> "Four elements are necessary to sustain a negligence claim: 1. A duty, or obligation, recognized by the law, requiring the [defendant] to conform to a certain standard of conduct, for the protection of others against unreasonable risks. 2. A failure on the [defendant's] part to conform to the standard required: a breach of the duty . . . . 3. A reasonably close causal connection between he conduct and the resulting injury. This is what is commonly known as 'legal cause,' or 'proximate cause,' and which includes the notion of cause in fact. 4. Actual loss or damage.

Curd v. Mosaic Fertilizer, LLC, ___ So.2d ___, 2010 WL 2400384 at *9 (Fla. 2010). Epstein does not allege a particular duty on the part of Edwards that has been breached. Nor does Epstein explain how any breach of the duty might have proximately caused him actual damages. Summary judgment is therefore appropriate for these reasons as well.

Finally, for the sake of completeness, it is worth noting briefly that no reasonable jury could find Edwards to have been negligent in failing to anticipate that a managing partner at his law firm would be involved in an unprecedented Ponzi scheme. Scott Rothstein deceived not only Edwards but also more than 60 other reputable lawyers at a major law firm. *Cf*. *Sun Sentinel, Fort* Lauderdale, Dec. 11, 2009, 2009 WLNR 25074193 at *1 ("Sure, some outlandish John Grisham murder plot[s] sound far-fetched. But if you asked me a few months ago if Scott Rothstein was fabricating federal court orders and forging a judge's signature on documents to allegedly fleece his friends, as federal prosecutors allege, I would have said that was far-fetched, too."). No reasonable lawyer could have expected that a fellow member of the bar would have been involved in such a plot. Nobody seemed to know of Rothstein's Ponzi scheme, not even his best friends, or the people he did business with on a daily basis, or even his wife. Many of the attorneys at RRA had been there for years and knew nothing. Edwards was a lawyer at RRA for less than 8 months and had very few personal encounters with Rothstein during his time at the firm, yet Epstein claims that he should have known of Rothstein's intricate Ponzi scheme. No doubt for this reason the U.S. Attorney's Office has now listed Edwards as a "victim" of Rothstein's crimes. *See* Statement of Undisputed Facts filed contemporaneously.

Epstein's Complaint does not offer any specific reason why a jury would conclude that Edwards was negligent, and he chose not to offer any explanation of his claim at his deposition. Accordingly, Edwards is entitled to summary judgment to the extent the claim against him is somehow dependent upon his negligence in failing to discover Rothstein's Ponzi scheme.

2. Edwards is Entitled to Summary Judgment to the Extent the Claim Against Him is Dependent on Allegations Regarding "Pumping the Cases" Because He Was Properly Pursuing the Interests of His Three Clients Who Had Been Sexually Abused by Epstein.

Epstein alleges that Edwards somehow improperly enhanced the value of the three civil cases he had filed against Epstein. Edwards represented three young women – L.M., E.W., and Jane Doe – by filing civil suits against Epstein for his sexual abuse of them while they were minors. Epstein purports to find a cause of action for this by alleging that Edwards somehow was involved in "'pumping' these three cases to investors."

As just explained, to the extent that Epstein is alleging that Edwards somehow did something related to the Ponzi scheme, those allegations fail for the simple reason that Edwards was not involved in any such scheme. Edwards, for example, could not have possibly "pumped" the cases to investors when he never participated in any communication with investors.

Epstein's "pumping" claims, however, fail for an even more basic reason: Edwards was entitled – indeed ethically obligated as an attorney – to secure the maximum recovery for his clients during the course of his legal representation. As is well known, "[a]s an advocate, a lawyer zealously asserts the client's position under the rules of the adversary system." Fla. Rules of Prof. Conduct, Preamble. Edwards therefore was required to pursue (unless otherwise instructed by his clients) a maximum recovery against Epstein. Edwards, therefore, cannot be liable for doing something that his ethical duties as an attorney required.[2]

Another reason that Epstein's claims that Edwards was "pumping" cases for investors fails is that Edwards filed all three cases almost a year before he was hired by RRA or even knew of Scott Rothstein. Epstein makes allegations that the complaints contained sensational allegations for the purposes of luring investors; however, language in the complaints remained virtually unchanged from the first filing in 2008 and from the overwhelming evidence the Court can see for itself that all of the facts alleged by Edwards in the complaints were true.

Epstein ultimately paid to settle all three of the cases Edwards filed against him for more money than he paid to settle any of the other claims against him. At Epstein's request, the terms of the settlement were kept confidential. The sum that he paid to settle all these cases is therefore not filed with this pleading and will be provided to the court for in-camera review. Epstein chose to make this payment as the result of a federal court ordered mediation process, which he himself sought (over the objection of Jane Doe, Edwards' client in federal court) in an effort to resolve the case. *See* Defendant's Motion for Settlement Conference, or in the Alternative, Motion to Direct Parties back to Mediation, *Doe v. Epstein*, No. 9:08-CV-80893 (S.D. Fla. June 28, 2010) (Marra, J.) (doc. #168) attached hereto as Exhibit "A". Notably, Epstein sought this settlement conference – and ultimately made his payments as a result of that conference - in July 2010, more than seven months after he filed this lawsuit against Edwards. Accordingly, Epstein could not have been the victim of any scheme to "pump" the cases against him, because he never paid to settle the cases until well after Edwards had left RRA and had severed all connection with Scott Rothstein (December 2009).

---

[2] In a further effort to harass Edwards, Epstein also filed a bar complaint with the Florida Bar against Edwards. The Florida Bar has dismissed that complaint. *See* Statement of Undisputed Facts.

In addition, if Epstein had thought that there was some improper coercion involved in, for example, Jane Doe's case, his remedy was to raise the matter before Federal District Court Judge Kenneth A. Marra who was presiding over the matter. Far from raising any such claim, Epstein simply chose to settle that case. He is therefore now barred by the doctrine of res judicata from somehow re-litigating what happened in (for example) the Jane Doe case. "The doctrine of res judicata makes a judgment on the merits conclusive 'not only as to every matter which was offered and received to sustain or defeat the claim, but as to every other matter which might with propriety have been litigated and determined in that action." *AMEC Civil, LLC v. State Dept. of Transp.*, ___ So.2d ___, 2010 WL 1542634 at *2 (Fla. 1st Dist. Ct. App. 2010) (*quoting Kimbrell v. Paige*, 448 So.2d 1009, 1012 (Fla. 1984). Obviously, any question of improper "pumping" of a particular case could have been resolved *in that very case* rather than now re-litigated in satellite litigation.

3. Edwards is Entitled to Summary Judgment on the Claim of Abuse of Process Because He Acted Properly Within the Boundaries of the Law in Pursuit of the Legitimate Interests of his Clients.

Epstein's Second Amended Complaint raises several claims of "abuse of process." An abuse of process claim requires proof of three elements: "(1) that the defendant made an illegal, improper, or perverted use of process; (2) that the defendant had ulterior motives or purposes in exercising such illegal, improper, or perverted use of process; and (3) that, as a result of such action on the part of the defendant, the plaintiff suffered damage." *S & I Investments v. Payless Flea Market, Inc.*, 36 So.3d 909, 917 (Fla. 4th Dist. Ct. App. 2010) (internal citation omitted). In fact, this Court is very familiar with this cause of action, as Edwards has correctly stated this cause in his counterclaim against Epstein. Edwards is entitled to summary judgment because Epstein cannot prove these elements.

The first element of an abuse of process claim is that a defendant made "an illegal, improper, or perverted use of process." On the surface, Epstein's Complaint appears to contain several allegations of such improper process. On examination, however, each of these allegations amounts to nothing other than a claim that Epstein was unhappy with some discovery proceeding, motion or argument made by Edwards. This is not the stuff of an abuse of process claim, particularly where Epstein fails to allege that he was required to do something as the result of Edwards' pursuit of the claims against him. *See Marty v. Gresh*, 501 So.2d 87, 90 (Fla. 1st Dist. Ct. App. 1987) (affirming summary judgment on an abuse of process claim where "appellant's lawsuit caused appellee to do nothing against her will").

In any event, none of the allegations of "improper" process can survive summary judgment scrutiny, because every action Edwards took was entirely proper and reasonably calculated to lead to the successful prosecution of the pending claims against Epstein as detailed in Edwards' Affidavit.

Epstein also fails to meet the second element of an abuse of process claim: that Edwards had some sort of ulterior motive. The case law is clear that on an abuse of process claim a "plaintiff must prove that the process was used for an immediate purpose other than that for which it was designed." *S&I Investments v. Payless Flea Market, Inc.*, 36 So.3d 909, 917 (Fla. 4th Dist. Ct. App. 2010) (*citing Biondo v. Powers*, 805 So.2d 67, 69 (Fla. 4th Dist. Ct. App. 2002). As a consequence, "[w]here the process was used to accomplish the result for which it was intended, regardless of an incidental or concurrent motive of spite or ulterior purpose, there is no abuse of process." *Id.* (internal quotation omitted). Here, Edwards has fully denied any improper motive, *See* Statement of Undisputed Facts, and Epstein has no evidence of any such motivation. Indeed, it is revealing that Epstein chose not to ask even a single question about this subject during the deposition of Edwards. In addition, all of the actions that Epstein complains about were in fact used for the immediate purpose of furthering the lawsuits filed by L.M., E.W., and Jane Doe. In other words, these actions all were both intended to accomplish and, in fact, successfully "accomplished the results for which they were intended" -- whether it was securing additional discovery or presenting a legal issue to the court handling the case or ultimately maximizing the recovery of damages from Epstein on behalf of his victims. Accordingly, Edwards is entitled to summary judgment on any claim that he abused process for this reason as well.

4. Edwards is Entitled to Summary Judgment to the Extent His Claim is Based On Pursuit of Discovery Concerning Epstein's Friends Because All Such Efforts Were Reasonably Calculated to Lead to Relevant and Admissible Testimony About Epstein's Abuse of Minor Girls.

Epstein has also alleged that Edwards improperly pursued discovery from some his close friends. Such discovery, Epstein claims, was improper because Edwards knew that these individuals lacked any discoverable information about the sexual assault cases against Epstein.

Here again, Edwards is entitled to summary judgment, as each of the friends of Epstein were reasonably believed to possess discoverable information. The undisputed facts show the following with regard to each of the persons raised in Epstein's complaint:

- With regard to Donald Trump, Edwards had sound legal basis for believing Mr. Trump had relevant and discoverable information. *See* Statement of Undisputed Facts.

- With regard to Alan Dershowitz (Harvard Law Professor), Edwards had sound legal basis for believing Mr. Dershowitz had relevant and discoverable information. *See* Statement of Undisputed Facts.

- With regard to former President Bill Clinton, Edwards had sound legal basis for believing former President Clinton had relevant and discoverable information. *See* Statement of Undisputed Facts.

- With regard to former Sony Record executive Tommy Mottola, Edwards was not the attorney that noticed Mr. Mottola's deposition. *See* Statement of Undisputed Facts.

- With regard to illusionist David Copperfield, Edwards had sound legal basis for believing Mr. Copperfield had relevant and discoverable information. *See* Statement of Undisputed Facts.

- With regard to former New Mexico Governor Bill Richardson, Edwards had sound legal basis for naming Former New Mexico Governor Bill Richardson on his witness list. *See* Statement of Undisputed Facts.

It is worth noting that the standard for discovery is a very liberal one. To notice someone for a deposition, of course, it is not required that the person deposed actually end up producing admissible evidence. Otherwise, every deposition that turned out to be a false alarm would lead to an "abuse of process" claim. Moreover, the rules of discovery themselves provide that a deposition need only be "reasonably calculated to *lead to* the discovery of admissible evidence." Fla. R. Civ. P. 1.280(b) (emphasis added).

Moreover, the discovery that Edwards pursued has to be considered against the backdrop of Epstein's obstructionist tactics. As the Court is aware, in both this case and all other cases filed against him, Epstein has asserted his Fifth Amendment privilege rather than answer any substantive questions. Epstein has also helped secure attorneys for his other household staff who assisted in the process of recruiting the minor girls, who in turn also asserted their Fifth Amendment rights rather than explain what happened behind closed doors in Epstein's mansion in West Palm Beach. *See* Statement of Undisputed Facts. It is against this backdrop that Edwards followed up on one of the only remaining lines of inquiry open to him: discovery aimed at Epstein's friends who might have been in a position to corroborate the fact that Epstein was sexually abusing young girls.

In the context of the sexual assault cases that Edwards had filed against Epstein, any act of sexual abuse had undeniable relevance to the case – even acts of abuse Epstein committed against minor girls other than L.M., E.W., or Jane Doe. Both federal and state evidence rules make acts of child abuse against other girls admissible in the plaintiff's case in chief as proof of "modus operandi" or "motive" or "common scheme or plan." *See* Fed. R. Evid. 415 (evidence of other acts of sexual abuse automatically admissible in a civil case); Fla. Stat. Ann. 90.404(b) (evidence of common scheme admissible); *Williams v. State*, 110 So.2d 654 (Fla. 1959) (other acts of potential sexual misconduct admissible).

A second reason exists for making discovery of Epstein's acts of abuse of other minor girls admissible. Juries considering punitive damages issues are plainly entitled to consider "the existence and frequency of similar past conduct." *TXO Production Corp. v. Alliance Resources Corp.*, 509 U.S. 443, 462 n.28 (1993). This is because the Supreme Court recognizes "that a recidivist may be punished more severely than a first offender . . . [because] repeated misconduct is more reprehensible than an individual instance of malfeasance." *BMW of North America, Inc. v. Gore*, 517 U.S. 559, 577 (1996) (supporting citations omitted). In addition, juries can consider other similar acts evidence as part of the deterrence calculation in awarding punitive damages, because "evidence that a defendant has repeatedly engaged in prohibited conduct while knowing . . . that it was unlawful would provide relevant support for an argument that strong medicine is required to cure the defendant's disrespect for the law." *Id.* at 576-77. In the cases Edwards filed against Epstein, his clients were entitled to attempt to prove that Epstein "repeatedly engaged in prohibited conduct" – i.e., because he was a predatory pedophile, he sexually assaulted dozens and dozens of minor girls. The discovery of Epstein's friends who might have had direct or circumstantial evidence of other acts of sexual assault was accordingly entirely proper. Edwards is therefore entitled summary judgment to the extent his claim is based on efforts by Edwards to obtain discovery of Epstein's friends.

III. EPSTEIN'S LAWSUIT MUST BE DISMISSED BECAUSE OF HIS REFUSAL TO PARTICIPATE IN REASONABLE DISCOVERY.

As is readily apparent from the facts of this case, Epstein has filed a lawsuit but then refused to allow any real discovery about the merits of his case. Instead, when asked hard questions about whether he has any legitimate claim at all, Epstein has hidden behind the Fifth

Amendment. As a result, under the "sword and shield doctrine" widely recognized in Florida caselaw, his suit must be dismissed.

"[T]he law is well settled that a plaintiff is not entitled to both his silence and his lawsuit." *Boys & Girls Clubs of Marion County, Inc. v. J.A.,* 22 So.3d 855, 856 (Fla. 5th Dist. Ct. App. 2009) (Griffin, J., concurring specially). Thus, "a person may not seek affirmative relief in a civil action and then invoke the fifth amendment to avoid giving discovery, using the fifth amendment as both a 'sword and a shield.'" *DePalma v. DePalma*, 538 So.2d 1290, 1290 (Fla. 4th Dist. Ct. App. 1989) (*quoting DeLisi v. Bankers Insurance Co.*, 436 So.2d 1099 (Fla. 4th Dist. Ct. App. 1983)). Put another way, "[a] civil litigant's fifth amendment right to avoid self-incrimination may be used as a shield but not a sword. This means that a plaintiff seeking affirmative relief in a civil action may not invoke the fifth amendment and refuse to comply with the defendant's discovery requests, thereby thwarting the defendant's defenses." *Rollins Burdick Hunter of New York, Inc. v. Euroclassic Limited, Inc.*, 502 So. 2d 959 (Fla. 3rd Dist. Court App. 1983).

Here, Epstein is trying to do precisely what the "well settled" law forbids. Specifically, he is trying to obtain "affirmative relief" – i.e., forcing Edwards to pay money damages – while simultaneously precluding Edwards from obtaining legitimate discovery at the heart of the allegations that form the basis for the relief Epstein is seeking. As recounted more fully in the statement of undisputed facts, Epstein has refused to answer such basic questions about his lawsuit as:

- "Specifically what are the allegations against you which you contend Mr. Edwards ginned up?"
- "Well, which of Mr. Edwards' cases do you contend were fabricated?"

- "Is there anything in L.M.'s Complaint that was filed against you in September of 2008 which you contend to be false?"
- "I would like to know whether you ever had any physical contact with the person referred to as Jane Doe in that [federal] complaint?"
- "Did you ever have any physical contact with E.W.?"
- "What is the actual value that you contend the claim of E.W. against you has?"

The matters addressed in these questions are the central focus of Epstein's claims against Edwards. Epstein's refusal to answer these and literally every other substantive question put to him in discovery has deprived Edwards of even a basic understanding of the evidence alleged to support claims against him. Moreover, by not offering any explanation of his allegations, Epstein is depriving Edwards of any opportunity to conduct third party discovery and opportunity to challenge Epstein's allegations.

It is the clear law that "the chief purpose of our discovery rules is to assist the truth-finding function of our justice system and to avoid trial by surprise or ambush," *Scipio v. State,* 928 So.2d 1138 (Fla.2006), and "full and fair discovery is essential to these important goals," *McFadden v. State*, 15 So.3d 755, 757 (Fla. 4th Dist. Ct. App. 2009). Accordingly, it is important for the Court to insure "not only compliance with the technical provisions of the discovery rules, but also adherence to the purpose and spirit of those rules in both the criminal and civil context." *McFadden*, 15 So.3d at 757. Epstein has repeatedly blocked "full and fair discovery," requiring dismissal of his claim against Edwards.

IV. EDWARDS IS ENTITLED TO ADVERSE INFERENCES FROM EPSTEIN'S INVOCATION OF THE FIFTH AMENDMENT AND THEREFORE TO SUMMARY JUDGMENT ON EPSTEIN'S CLAIM.

Edwards is entitled to summary judgment on the claim against him for a second and entirely independent reason: Epstein's repeated invocations of the Fifth Amendment raise adverse inferences against him that leave no possibility that a reasonable factfinder could reach a verdict in his favor. In ruling on a summary judgment motion, the court must fulfill a "gatekeeping function" and should ask whether "a *reasonable* trier of fact could possibly" reach a verdict in favor of the plaintiff. *Willingham v. City of Orlando*, 929 So.2d 43, 48 (Fla. 5th Dist. Ct. App. 2006) (emphasis added). Given all of the inferences that are to be drawn against Epstein, no reasonable finder of fact could conclude that Epstein was somehow the victim of improper civil lawsuits filed against him. Instead, a reasonable finder of fact could only find that Epstein was a serial molester of children who was being held accountable through legitimate suits brought by Edwards and others on behalf of the minor girls that Epstein victimized.

"[I]t is well-settled that the Fifth Amendment does not forbid adverse inferences against parties to civil actions when they refuse to testify in response to probative evidence offered against them." *Baxter v. Palmigiano,* 425 U.S. 308, 318 (1976); *accord Vasquez v. State*, 777 So.2d 1200, 1203 (Fla. App. 2001). The reason for this rule "is both logical and utilitarian. A party may not trample upon the rights of others and then escape the consequences by invoking a constitutional privilege – at least not in a civil setting." *Fraser v. Security and Inv. Corp.*, 615 So.2d 841, 842 (Fla. 4th Dist. Ct. App. 1993). And, in the proper circumstances, "'Silence is often evidence of the most persuasive character.'" *Fraser v. Security and Inv. Corp.*, 615 So.2d

841, 842 (Fla. 4th Dist. Ct. App. 1993) (*quoting United States ex rel. Bilokumsky v. Tod,* 263 U.S. 149, 153-154 (1923) (Brandeis, J.).

In the circumstances of this case, a reasonable finder of fact would have "evidence of the most persuasive character" from Epstein's repeated refusal to answer questions propounded to him. To provide but a few examples, here are questions that Epstein refused to answer and the reasonable inference that a reasonable finder of fact would draw:

- Question not answered: "Specifically what are the allegations against you which you contend Mr. Edwards ginned up?" Reasonable inference: No allegations against Epstein were ginned up.

- Question not answered: "Well, which of Mr. Edwards' cases do you contend were fabricated?" Reasonable inference: No cases filed by Edwards against Epstein were fabricated.

- Question not answered: "Did sexual assaults ever take place on a private airplane on which you were a passenger?" Reasonable inference: Epstein was on a private airplane while sexual assaults were taking place.

- Question not answered: "How many minors have you procured for prostitution?" Reasonable inference: Epstein has procured multiple minors for prostitution.

- Question not answered: "Is there anything in L.M.'s Complaint that was filed against you in September of 2008 which you contend to be false?" Reasonable inference: Nothing in L.M.'s complaint filed in September of 2008 was false – i.e., as alleged in L.M.'s complaint, Epstein repeatedly sexually assaulted her while she was a minor and she was entitled to substantial compensatory and punitive damages as a result.

- Question not answered: "I would like to know whether you ever had any physical contact with the person referred to as Jane Doe in that [federal] complaint?" Reasonable inference: Epstein had physical contact with minor Jane Doe as alleged in her federal complaint.

- Question not answered: "Did you ever have any physical contact with E.W.?" Reasonable inference: Epstein had physical contact with minor E.W. as alleged in her complaint.

- Question not answered: "What is the actual value that you contend the claim of

> E.W. against you has?" Reasonable inference: E.W.'s claim against Epstein had substantial actual value.

Without repeating each and every invocation of the Fifth Amendment that Epstein has made and the reasonable inferences to be drawn from those invocations of privilege, the big picture is unmistakably clear: No reasonable finder of fact could rule in Epstein's favor on his claims against Edwards. Accordingly, Edwards is entitled to summary judgment based on the Fifth Amendment inferences that the jury would draw.

The inferences against Epstein are not limited to those arising from his privilege assertions. Epstein's guilt is also reasonably inferred from his harassment of, intimidation of, efforts to exercise control over, and limitation of access to witnesses who might testify against him.

Epstein's efforts to intimidate his victims support the inference that Epstein knew that they were going to provide compelling testimony against him. The evidence that Epstein tampered with witnesses (later designated as his accomplices and co-conspirators) will be admissible to demonstrate his consciousness of guilt. "[I]t is precisely because of the egregious nature of such conduct that the law expressly permits the jury to make adverse inferences from a party's efforts to intimidate witnesses . . . ." *Jost v. Ahmad*, 730 So.2d 708, 711 (Fla. 2nd Dist. Ct. App. 1998) (internal quotation omitted). To be clear, Epstein's attempt to tamper with witnesses is "not simply admissible as impeachment evidence of the tampering party's credibility. The opposing party is entitled to introduce facts regarding efforts to intimidate a witness as *substantive evidence*." *Id.* at 711 (emphasis in original) (internal citation omitted). This substantive evidence of Epstein's witness intimidation provides yet another reason why no reasonable jury could find in favor of his claims against Edwards.

V. EDWARDS IS ENTITLED TO SUMMARY JUDGMENT ON THE BASIS OF HIS AFFIRMATIVE DEFENSE OF PRIVILEGE

Absolute immunity must be afforded any act occurring during course of judicial proceeding, regardless of whether act involves defamatory statement or other tortious behavior, such as tortious interference with business relationship, so long as act has some relationship to proceeding. *See Levin, Middlebrooks, Mabie, Thomas, Mayes & Mitchell, P.A. v. U.S. Fire Ins. Co.*, 639 So. 2d 606 (Fla. 1994). The immunity afforded to statements made during the course of a judicial proceeding extends not only to the parties in a proceeding but to judges, witnesses, and counsel as well. *Id.* The litigation privilege applies in all causes of action, whether for common-law torts or statutory violations. *See Echevarria, McCalla, Raymer, Barrett & Frappier v. Cole*, 950 So. 2d 380 (Fla. 2007). Defamatory statements made by lawyer while interviewing a witness in preparation for and connected to pending litigation are covered by the absolute immunity conferred by the litigation privilege. *See DelMonico v. Traynor*, 50 So. 3d 4 (Fla. Dist. Ct. App. 4th Dist. 2010), review granted, 47 So. 3d 1287 (Fla. 2010). The privilege extends to statements in judicial proceedings or those "necessarily preliminary thereto. *See Stewart v. Sun Sentinel Co.,* 695 So.2d 360 (Fla. 4th DCA 1997)(an attorney's delivery of a copy of a notice of claim to a reporter, which notice was a required filing prior to instituting suit, was protected by absolute immunity).

CONCLUSION

For all the foregoing reasons, defendant, the Court should grant defendant Bradley J. Edwards, Esq., summary judgment in his favor on the only remaining claim filed against him by plaintiff Jeffrey Epstein, and any other relief that the Court deems just and proper.

CERTIFICATE OF SERVICE

I HEREBY CERTIFY that on November 3rd, 2011 a copy of the foregoing has been served via Fax and U.S. Mail to all those on the attached service list.

Jack Scarola
Searcy, Denney, Scarola, Barnhart & Shipley
2139 Palm Beach Lakes Blvd
West Palm Beach, FL 33409
(561) 686-6300
(561) 684-5816 (fax)

By: ______________________
JACK SCAROLA
Florida Bar No.: 169440

SERVICE LIST

Christopher E. Knight, Esq.
Joseph L. Ackerman, Esq.
FOWLER WHITE BURNETT P.A.
901 Phillips Point West
777 South Flagler Drive
West Palm Beach, FL 33401

Jack Alan Goldberger, Esq.
Atterbury Goldberger et al.
250 Australian Avenue South
Suite 1400
West Palm Beach, FL 33401

Marc S. Nurik, Esq.
Law Offices of Marc S. Nurik
One E. Broward Blvd., Suite 700
Fort Lauderdale, FL 33301

Gary M. Farmer, Jr.
Farmer, Jaffe, Weissing,
Edwards, Fistos & Lehrman, P.L.
425 N. Andrews Ave., Suite 2
Fort Lauderdale, FL 33301

IN THE CIRCUIT COURT OF THE
FIFTEENTH JUDICIAL CIRCUIT, IN AND
FOR PALM BEACH COUNTY, FLORIDA

CASE NO.: 502009CA040800XXXXMBAG

JEFFREY EPSTEIN,

Plaintiff(s),

vs.

SCOTT ROTHSTEIN, individually,
BRADLEY J. EDWARDS, individually, and
L.M., individually,

Defendant(s).
_______________________________________/

COUNTER-PLAINTIFF, EDWARDS' SECOND RENEWED MOTION FOR LEAVE TO ASSERT CLAIM FOR PUNITIVE DAMAGES

Counter-plaintiff, BRADLEY J. EDWARDS, moves this Honorable Court for entry of an Order granting him leave to assert a claim for punitive damages against the Counter-defendant, JEFFREY EPSTEIN, and in support thereof would show that the evidence summarized herein satisfies the statutory prerequisites for the assertion of a punitive damage claim. Specifically, the evidence establishes that EPSTEIN's Complaint against EDWARDS;

1. was filed in the total absence of evidence to support any allegation of wrongdoing on the part of EDWARDS;
2. was filed in the total absence of evidence that EPSTEIN had sustained damage as a consequence of any misconduct other than his own well-established criminal enterprise;
3. was filed in the absence of any intention to meet his own obligation to provide relevant and material discovery;

EXHIBIT
C

4. was filed for the sole purpose of attempting to intimidate both EDWARDS and EDWARDS' clients and others into abandoning their legitimate claims against EPSTEIN.

APPLICABLE LAW

To plead a claim for punitive damages, the claimant must show a "reasonable basis" for the recovery of such damages. *See* Fla.R.Civ.P. 1.190(f); *see also Globe Newspaper Co. v. King,* 658 So.2d 518, 520 (Fla. 1995). The showing required to amend is minimal. As stated in *State of Wis. Inv. v. Plantation Square Assoc.,* 761 F. Supp. 1569, 1580 (S.D. Fla. 1991):

> [T] he court believes it must ultimately be a lesser standard than that required for summary judgment. Though the burden is on [the plaintiff] to survive a §768.72 challenge of insufficiency, *see Will v. Systems Engineering Consultants,* 554 So.2d 591, 592 (Fla. 3rd DCA 1989), the standard of proof required to assert Plaintiff's punitive claim must be lower than that needed to survive a summary adjudication on its merits. As the Florida courts have noted, a §768.72 challenge more closely resembles a motion to dismiss that additionally requires an evidentiary proffer and places the burden of persuasion on the plaintiff. *Id.* In considering a motion to dismiss, factual adjudication is inappropriate as all facts asserted—or here, reasonably established—by the plaintiff are to be taken as true. *Conley v. Gibson,* 355 U.S. 41, at 45-46, 78 S. Ct. 99, at 101-102, 2 L.Ed. 2d 80, 1581 at 84. As such, the court has given recognition only to those assertions of the defendants which would show Plaintiff's factual bases to be patently false or irrelevant, and has paid no heed whatsoever to the defendants' alternative evidentiary proffers.

State of Wis. Inv., 761 F. Supp. At 1580; *see also Dolphin Cove Assn. v. Square D. Co.,* 616 So. 2d 553 (Fla. 2d DCA 1993) ("Prejudging the evidence is not a proper vehicle for the court's denial of the motion to amend" to assert punitive damages claim).

Section 768.72 provides for the amendment of a complaint either through evidence in the record *or* "proffered by the claimant." As the statute suggests, a proffer of evidence in support of a punitive damage claim is sufficient and a formal evidentiary hearing is not required. *See Strasser v. Yalmanchi,* 677 So.2d 22, 23 (Fla. 4th DCA 1996), *rev. dismissed,* 699 So.2d 1372 (Fla. 1997); *Solis v. Calvo,* 689 So.2d 366, 369, n.2 (Fla. 3d DCA 1997). In fact, a hearing is not even required provided the trial court identifies the filings of the parties and indicates that its decision to grant the motion is based upon a review of the file and the respective documents filed.

The United States District Court for the Middle District of Florida has spoken clearly on the nature of a proffer in support of a motion to amend to assert a claim for punitive damages in *Royal Marco Point I Condo. Ass'n, Inc. v. QBE Ins. Corp.*, 2010 WL 2609367 (M.D. Fla. June 30, 2010). As the Court stated:

> It is important to emphasize, at the outset, the limited nature of the review a court may undertake in considering the sufficiency of an evidentiary proffer under Fla. Stat. §768.72. Courts reviewing such proffers have recognized that "a 'proffer' according to traditional notions of the term, connotes merely an 'offer' of evidence and neither the term standing alone nor the statute itself calls for an adjudication of the underlying veracity of that which is submitted, much less for countervailing evidentiary submissions." *Estate of Despain v. Avante Group, Inc.,* 900 So.2d 637, 642 (Fla. 5th DCA 2005) (quoting *State of Wisconsin Investment Board v. Plantation Square Associates, Ltd.,* 761 F. Supp. 1569, 1581 n. 21 (S.D. Fla. 1991)).
>
> Therefore, "an evidentiary hearing where witnesses testify and evidence is offered and scrutinized under the pertinent evidentiary rules, as in a trial, is neither contemplated nor mandated by the statute in order to determine whether a reasonable basis has been established to plead punitive damages." *Id.* (collecting cases).

> It is thus neither necessary nor appropriate for a court to make evidentiary rulings, weigh rebuttal evidence, or engage in credibility determinations in considering the sufficiency of the proffer.
>
> "…a proffer should be evaluated by standards akin to those governing a motion to dismiss, where the truth of the plaintiff's allegations are assumed, and not the more rigorous summary judgment standard, where the opposing party must show that there is sufficient admissible evidence in the record to support a reasonable jury finding in his favor."

I. INTRODUCTION

The pleadings and discovery taken to date as confirmed by Epstein's voluntary dismissal of all claims brought by him against Bradley J. Edwards, show that there is an absence of competent evidence to demonstrate that Edwards participated in any fraud against Epstein, show the propriety of every aspect of Edwards' involvement in the prosecution of legitimate claims against Epstein, and further support the conclusion that Epstein sued Edwards out of malice and for the purpose of intending to intimidate Edwards and Edwards' clients into abandoning or compromising their legitimate claims against Epstein. Epstein sexually abused three clients of Edwards – L.M., E.W., and Jane Doe – and Edwards properly and successfully represented them in a civil action against Epstein. Nothing in Edwards's capable and competent representation of his clients could serve as the basis for a civil lawsuit against him. Allegations about Edwards's participation in or knowledge of the use of the civil actions against Epstein in a "Ponzi Scheme" were never supported by probable cause or any competent evidence and could never be supported by competent evidence as they are entirely false.

A. Epstein's Complaint

Epstein's Second Amended Complaint essentially alleged that Epstein was damaged by Edwards, acting in concert with Scott Rothstein (President of the Rothstein Rosenfeldt Adler law firm ("RRA") where Edwards worked for a short period of time). Epstein appeared to allege that Edwards joined Rothstein in the abusive prosecution of sexual assault cases against Epstein to "pump" the cases to Ponzi scheme investors. As described by Epstein, investor victims were told by Rothstein that three minor girls who were sexually assaulted by Epstein: L.M., E.W., and Jane Doe were to be paid up-front money to prevent those girls from settling their civil cases against Epstein. In Epstein's view, these child sexual assault cases had "minimal value" (Complaint & 42(h)), and Edwards's refusal to force his clients to accept modest settlement offers was claimed to breach some duty that Edwards owed to Epstein. Interestingly, Epstein never states that he actually made any settlement offers.

The supposed "proof" of the Complaint's allegations against Edwards includes Edwards's alleged contacts with the media, his attempts to obtain discovery from high-profile persons with whom Epstein socialized, and use of "ridiculously inflammatory" language in arguments in court. Remarkably, Epstein has filed such allegations against Edwards despite the fact that Epstein had sexually abused each of Edwards's clients and others while they were minors. Indeed, in discovery Epstein has asserted his Fifth Amendment privilege rather than answer questions about the extent of the sexual abuse of his many victims. Even more remarkably, since filing his suit against Edwards, Epstein settled the three cases Edwards handled for an amount that Epstein insisted be kept confidential. Without violating the strict confidentiality terms required by Epstein, the cases did not settle for the "minimal value" that Epstein suggested in his Complaint. Because Epstein relies upon the alleged discrepancy between the "minimal value" Epstein ascribed to the claims and the substantial value Edwards sought to recover for his clients, the settlement amounts Epstein voluntarily agreed to pay while these claims against Edwards were pending will be disclosed to the court in-camera.

B. Summary of the Argument

The claims against Bradley J. Edwards, Esq., were frivolous for at least three separate reasons.

First, because Epstein elected to hide behind the shield of his right against self-incrimination to preclude his disclosing any relevant information about the criminal activity at the center of his claims, he was barred from prosecuting his case against Edwards. Under the well-established "sword and shield" doctrine, Epstein could not legitimately seek damages from Edwards while at the same time asserting a Fifth Amendment privilege to block relevant discovery. His case was therefore subject to summary judgment and on the eve of the hearing seeking that summary judgment Epstein effectively conceded that fact by voluntarily dismissing his claims.

Second, all of Edwards' conduct in the prosecution of valid claims against Epstein was protected by the litigation privilege, a second absolute legal bar to Epstein's claims effectively conceded by his voluntary dismissal.

Third, and most fundamentally, Epstein's lawsuit was not only unsupported by both the applicable law, it was based on unsupported factual allegations directly contradicted by all of the record evidence. From the beginning, Edwards diligently represented three victims of sexual assaults perpetrated by Epstein. As explained in detail below, each and every one of Edwards's litigation decisions was grounded in proper litigation judgment about the need to pursue effective discovery against Epstein, particularly in the face of Epstein's stonewalling tactics. Edwards's successful representation finally forced Epstein to settle and pay appropriate damages. Effective and proper representation of child victims who have been repeatedly sexually assaulted cannot form the basis of a separate, "satellite" lawsuit, and therefore Edwards is entitled to summary judgment on these grounds as well.

The truth is the record is entirely devoid of any evidence to support Epstein's claims and is completely and consistently corroborative of Edwards's sworn assertion of innocence. Put simply, Epstein made allegations that have no basis in fact. To the contrary, his lawsuit was merely a desperate measure by a serial pedophile to prevent being held accountable for repeatedly sexually abusing minor females. Epstein's ulterior motives in filing and prosecuting this lawsuit are blatantly obvious. Epstein's behavior is another clear demonstration that he feels he lives above the law and that because of his wealth he can manipulate the system and pay for lawyers to do his dirty work - even to the extent of having them assert baseless claims against other members of the Florida Bar. Every one of Epstein's Complaints against Edwards was nothing short of a far-fetched fictional fairy-tale with absolutely no evidence whatsoever to support his preposterous claims. It was his last ditch effort to escape the public disclosure by Edwards and his clients of the nature, extent, and sordid details of Epstein's life as a serial child molester.

ARGUMENT

II. THE RECORD AND PROFFERED EVIDENCE ESTABLISHES THAT EDWARDS'S CONDUCT COULD NOT POSSIBLY FORM THE BASIS OF ANY LIABILITY IN FAVOR OF EPSTEIN

A. The Summary Judgment Standard.

Rule 1.510(c), Florida Rules of Civil Procedure, provides that a court may enter summary judgment when the pleadings, depositions and factual showings reveal that there is no genuine issue of material fact and that the moving party is entitled to judgment as a matter of law. *See Snyder v. Cheezem Development Corp.*, 373 So. 2d 719, 720 (Fla. 2d DCA 1979); Rule 1.510(c), Fla. R. Civ. P. Once the moving party conclusively establishes that the nonmoving party cannot prevail, it is incumbent on the nonmoving party to submit evidence to rebut the motion for summary judgment. *See Holl v. Talcott*, 191 So. 2d 40, 43 (Fla. 1966). It is not enough for the opposing party merely to assert that an issue of fact does exist. *Fisel v. Wynns*, 667 So.2d 761, 764 (Fla.1996); *Landers v. Milton,* 370 So.2d 368, 370 (Fla.1979) (same).

Moreover, it is well-recognized that the non-moving party faced with a summary judgment motion supported by appropriate proof may not rely on bare, conclusory assertions found in the pleadings to create an issue and thus avoid summary judgment. Instead, the party must produce counter-evidence establishing a genuine issue of material fact. *See Bryant v. Shands Teaching Hospital and Clinics, Inc.,* 479 So.2d 165, 168 (Fla. 1st Dist. Ct. App. 1985); *see also Lanzner v. City of North Miami Beach,* 141 So.2d 626 (Fla. 3d Dist Ct. App. 1962) (recognizing that mere contrary allegations of complaint were not sufficient to preclude summary judgment on basis of facts established without dispute). Where the nonmoving party fails to present evidence rebutting the motion for summary judgment and there is no genuine issue of material fact, then entry of judgment is proper as a matter of law. *See Davis v. Hathaway*, 408 So. 2d 688, 689 (Fla. 2d Dist. Ct. App. 1982); *see also Holl*, 191 So. 2d at 43. Faced with these well-established legal principles, Epstein voluntarily dismissed his claims against Edwards on the eve of the hearing on Edwards Motion for Summary Judgment.

B. Epstein's Claim Regarding Edwards Had Absolutely No Factual Basis.

This was not a complicated case for granting summary judgment. To the contrary, the uncontested record clearly established that each and every one of Epstein's claims against Edwards lacked any merit whatsoever.[1]

1. Epstein's allegations regarding Edwards' involvement in Rothstein's "Ponzi Scheme" were unsupported and unsupportable because Edwards was simply not involved in any such scheme.

a. Edwards Had No Involvement in the Ponzi Scheme.

The bulk of Epstein's claims against Edwards hinged on the premise that Edwards was involved in a Ponzi scheme run by Scott Rothstein. Broad allegations of wrongdoing on the part of Edwards were scattered willy-nilly throughout the complaint. None of the allegations provided any substance as to how Edwards actually assisted the Ponzi scheme, and allegations that he "knew or should have known" of its existence are based upon an impermissible pyramiding of inferences. In any event, these allegations all fail for one straightforward reason:

---

[1] The dismissal of Epstein's claims against Edwards did not affect Epstein's claims against Scott Rothstein. Epstein had already chosen to dismiss all of his claims against L.M., the only other defendant named in the suit.

1 of 11 DOCUMENTS

Copyright 2009 Associated Newspapers Ltd.
All Rights Reserved

The Evening Standard (London)

December 24, 2009 Thursday

LENGTH: 824 words

HEADLINE: CITY SPY

BODY:

EXPECT more media firms to announce plans to charge for content online in early 2010. City Spy hears that business-to-business publisher United Business Media is the latest outfit which is thinking of ramping up its subscription model. Property Week and Building are among the titles which recently started asking users to register their details to keep reading stories, which is seen as a possible precursor to charging.

BUSINESSES TIPPED TO COME A CROPPER

AMID all the contradictory forecasts for recovery or double-dip recession in 2010, what do the insolvency practitioners say? City Spy's mole in the bean-counting world says the last quarter of 2009 was surprisingly quiet as the economy stabilised but they are not optimistic about the new year: "We reckon there's going to be a rush of insolvencies in the second quarter, after the end of the financial year." The next quarterly rent review is due tomorrow, Christmas Day, then again at the end of March. But given the number of "seasonal sales" that started on the High Street at least a week before Christmas, it would be no surprise to see some retailers come a cropper sooner...

EPSTEIN PILOT TAKES TO THE ROAD

FURTHER news reaches City Spy of former Bear Stearns trader, Prince Andrew's shooting companion and convicted sex offender, Jeffrey Epstein.

The ex-Wall Street star served 13 months in jail on criminal charges of soliciting prostitution and procuring a minor for prostitution and he now faces civil claims from young women accusing him of having unlawful sex with them. This week, City Spy recounted how Epstein had transferred the title deeds of his prized 2003 Ferrari 575M Maranello to his private pilot Larry Visoki, prior to the car going on sale for $159,000 (£99,000) (possibly to help Epstein pay his legal bills). It turns out, the same Visoki was deposed last week by Bradley Edwards, an attorney for three of the women suing Epstein. Questioned by Edwards about plane passengers who might have witnessed Epstein in the company of young girls, Visoki admitted Bill Clinton, Prince Andrew, former Israeli prime minister Ehud Barak, former Colombian president Andrés Pastrana Arango, Obama economic adviser Lawrence Summers, billionaire Ron Burkle, and actors Kevin Spacey and Chris Tucker had been on board the plane while young girls were present. Fortuitously for Epstein, however, Ferrari-selling Visoki swore on oath that he never suspected his boss of having sex with them.

Of course not, Larry. Now drive off into the sunset.

More on Prince Andrew, our special representative for international trade and investment. The European Parliament and the Organisation for Security and Co-operation in Europe have strongly condemned Azerbaijan for tightening restrictions on the media and jailing two bloggers who were critical of the government. It transpires the oil-rich country has long blocked BBC broadcasts there, which might explain why oft-criticised Andrew and former Prime Minister Tony Blair spend so much time visiting the sometime Soviet State.

What does the snow have in common with the recession? Every other country can get out of both but Britain can't get out of either. HAPPY news: private jet travel is back, reports the Wall Street Journal. Alas, there is a "but" [#x2039] in-flight food remains in recession. Apparently, those who supply food to executive aircraft are seeing demand soar after a slump, but says one caterer: "No one is eating lobster. A quick turkey box lunch is the order of the day." Of course, that has nothing to do with the industry being desperate to re-brand itself as time-saving and cost-efficient.

WHICH insurance broker saw a compliance officer pass out after the office Christmas lunch and have to be taken to hospital?

WHO MADE OFF WITH THE MONEY? IT's a year since the Bernie Madoff affair blew up and the hedge fund king was found to have been ripping off his clients. If he was in Britain the old fraudster would still be at liberty as lawyers pored over his case and the prosecution had barely cranked into operation. But the US is different [#x2039] his case is done and dusted, and he's languishing in jail. Even so, by US standards, the Madoff conviction was going some. Rumours persist that he pleaded guilty as quickly as he did and said the absolute minimum because he wasn't the main crook of the piece [#x2039] the main business of his hedge fund was washing money for organised crime. As soon as the balloon went up and he was arrested, he was warned by friends with Italian-American origins that his life, and the lives of his family, would be at risk were he not to "take the rap".

OETaking the rap': hedge fund fraud Bernie Madoff

UNFORTUNATE name? City Spy's eye is drawn to a forthcoming lecture at the Institute of Advanced Legal Studies, School of Advanced Study, University of London. It's in partnership with the Market Abuse Association. What? Do they wear a club tie? Do they refer to each other as fellow market abusers?

LOAD-DATE: December 24, 2009

From: Darren Indyke [redacted]
To: Jackie Perczek [redacted]
Date: 4/7/2011 1:44 PM
Subject: Privileged and Confidential
Attachments: Attorneys Say Miami Prosecutors Violated Crime Victims' Rights Act | Main Justice.pdf; Attorneys want Jeffrey Epstein agreement thrown out.pdf; Edwards Articles - Rush Interview 2.pdf; Edwards Articles - Rush Interview.pdf; Part.005

Additional Articles.

Darren K. Indyke
Darren K. Indyke, PLLC
301 East 66th Street, 10B
New York, New York 10065
Telephone: (212) 517-2052
Direct: (646) 862-4817
Fax: (212) 517-7779
email: dkiesq@aol.com

IEGOTIATING WITH THE DOJ:
STRATEGIES FOR OPTIMAL RESULTS
FREE Webinar - Wed., April 13
MAIN JUSTICE
BRACEWELL &GIULIANI
Wanted: Lawyers with DOJ Experience
MJobs

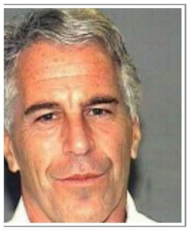
AIN JUSTICE
LITICS, POLICY AND THE LAW

orneys Say Miami Prosecutors Violated Crime Victims' Rights Act

tephanie Woodrow | March 22, 2011 11:52 am

Printable Version
Rights/Reprints

ıse attorneys for two girls who contend they were assaulted by billionaire and cted sex offender **Jeffrey Epstein** filed court papers on Monday claiming the ı U.S. Attorney's office violated the Crime Victims' Rights Act by signing a nonprosecution agreement with Epstein ut notifying them, the Palm Beach Daily News reported.



y Epstein (gov)

Epstein served 13 months in jail from June 2008 to July 2009 for one state count of soliciting an underage girl for prostitution. As a result, he is required to register as a sex offender. While he wasn't prosecuted for additional charges, more than 40 girls under the age of 18 say they came to his home and gave him massages. During the massages, they say, he masturbated and sexually assaulted them.

Brad Edwards and **Paul Cassell**, attorneys representing two of his alleged victims, say in the filing that the U.S. Attorney's office for the Southern District of Florida of deliberately misled the victims by telling them there was an ongoing investigation into their claims. However, they say, the office was concealing the fact that they already had signed a rosecution deal with Epstein.

rding to the motion, the U.S. Attorney's office in January 2008 and May 2008 sent "false notification" letters in to ein's alleged victims saying "(t)his case is currently under investigation." However, the office had signed the agreement Epstein in September 2007.

ıttorneys want a court hearing during which they will ask that the agreement be invalidated because it violated the ıs' rights. In the motion, the attorneys claim the agreement is illegal because the government did not protect the gressionally mandated rights of victims before it entered this agreement."

dge grants the request, Epstein could be charged by the U.S. Attorney's office. If he were charged and convicted on al charges, he could be sentenced to 10 years to life for each charge.

rding to the motion, "The only reason that the (U.S. Attorney's office) concealed the existence of the non-prosecution ement from the victims was not to comply with some legal restriction, but rather to avoid a firestorm of public oversy that would have erupted if the sweetheart plea deal with a politically connected billionaire had been revealed."

ı Valle, special counsel for the U.S. Attorney's Office Southern District of Florida, in an email to the newspaper that the Attorney's office will respond in court filings.

ever, as we stated more than two years ago in July 2008 in our response to the plaintiffs' then-emergency petition for cement of the Crime Victim Rights Act, the CVRA was not violated because no federal charges were ever filed in the ıern District of Florida," Valle said. "Because the matter remains pending in court, it would be inappropriate at this time ovide additional comment on the merits of the current motion."

ATED POSTS:

U.S. ATTORNEYS CHART (Interactive)

Track leadership changes in the 93 U.S. Attorney offices, with Bush holdovers and Obama candidates.

MILLER CHEVALIER
Providing clients with proven experience and innovative solutions to complex compliance and enforcement issues inside the Beltway and around the globe.
Miller & Chevalier Chartered
millerchevalier.com

COVINGTON
COVINGTON & BURLING LLP
Recognized for its Leading White Collar Crime and Anti-Corruption Practices by *Chambers* and *Legal 500*

Decisions about who, where and how to prosecute have always been--and must remain--the responsibility of the executive branch." - Attorney General Eric Holder discussing pressure from members of Congress to prosecute Khalid Sheikh Mohammed before a military commission rather than in *federal court.*

ATED POSTS:

- ɔama To Tap Non-Profit Head For DOJ Crime Victims Post
- wmakers Bemoan Lack of Funding For Victims' Rights
- osecutor Who Violated Rights of Islamic Groups by Accident is Now U.S. Attorney
- emocratic Donor Violated Spirit of Election Laws, LA Prosecutors Say
- deral Prosecutors Violated Laws, Ethics Rules in 201 Cases Since 1998, Study Finds

LDER POST
erez Asks States to Help Enforce ADA Protections for tudents

NEWER POST
Facing Hiring Freeze, Kansas U.S. Attorney Wants More Lawyers

nents are closed.

JUSTICE DEPARTMENT NEWS RELEASES

- **Associate Attorney General Tom Perrelli Speaks at the Department of Educationâ��s Gender-Based Violence Summit**
- **Comverse Technology INC. Agrees to Pay $1.2 Million Penalty to Resolve Violations of the Foreign Corrupt Practices Act**
- **Attorney General Eric Holder Speaks at National Action Networkâ��s 13th Annual Convention**
- **Shenandoah, Pennsylvania, Man Senteced for Involvement in the Fatal Beating of Luis Ramirez**
- **New Jersey Wastewater Treatment and Chemical Supply Company and Owner Sentenced for Their Role in Fraud Conspiracy**
- **Alabama Doctor and Husband Charged with Tax Evasion**
- **Tax Defendant Indicted for Filing False Liens for Billions of Dollars Against Federal Law Enforcement**
- **JGC Corporation Resolves Foreign Corrupt Practices Act Investigation and Agrees to Pay a $218.8 Million Criminal Penalty**
- **Attorney General Eric Holder Speaks at National Forum on Youth Violence Prevention Summit**
- **Five Individuals Indicted for Alleged Roles in Scheme to Defraud Program Providing Matching Funds Contributions to Non-Profit Organization**

Government Sites

Department of Justice
House Judiciary Committee
Office of Government Ethics
Office of Legal Counsel
Office of Professional Responsiblity
Senate Judiciary Committee

ne | About | Advertise | Jobs | Privacy Policy | Terms of Use

Copyright © 2011 WDC Media LLC

JUST A CLICK AWAY! Dining Directory PalmBeachDailyNews.com

Attorneys want Jeffrey Epstein agreement thrown out

LATEST NEWS

- Today: Thursday, April 7, 2011
- Talk at Sea Gull Cottage to focus on genomics
- Palm Beach Women's International Film Festival debuts Thursday in support of women filmmakers
- Day Academy pupils revisit notable characters in Palm Beach history
- Gov. Scott at Midtown Beach: 'Give me the list,' he tells officials about beach concerns

By **MICHELE DARGAN**
DAILY NEWS STAFF WRITER

Updated: 9:41 a.m. Wednesday, March 23, 2011
Posted: 7:21 p.m. Monday, March 21, 2011

E-mail | Print | Share | Larger Type

Court papers filed Monday say the U.S. Attorney's Office violated the Crime Victims' Rights Act by signing a nonprosecution agreement with sex offender Jeffrey Epstein without notifying his victims.

Attorneys Brad Edwards and Paul Cassell, representing Jane Doe #1 and Jane Doe #2, want a court hearing, where they will ask that the agreement be invalidated because, they say, the victims' rights were violated. If that happens, it could open up the 58-year-old Palm Beach billionaire to a slew of federal charges involving sex crimes with minors that were set aside by the agreement.

The motion, filed Monday in federal court in West Palm Beach, accuses the U.S. Attorney's Office of deliberately misleading the victims by telling them the investigation was ongoing, while concealing they had already signed a deal with Epstein.

According to the motion, the U.S. Attorney's Office sent "false notification" letters in January 2008 and May 2008 to the victims saying "(t)his case is currently under investigation" after the government had signed the agreement with Epstein in September 2007.

"The only reason that the (U.S. Attorney's Office) concealed the existence of the non-prosecution agreement from the victims was not to comply with some legal restriction, but rather to avoid a firestorm of public controversy that would have erupted if the sweetheart plea deal with a politically connected billionaire had been revealed," the motion says.

If Epstein were found guilty on federal charges, statutory penalties ranged from 10 years to life.

Instead, the sealed pact was part and parcel of Epstein's acceptance of a state plea deal, where he received an 18-month sentence for soliciting a minor for prostitution and soliciting prostitution. He served 13 months segregated in a vacant wing of the county stockade and was let out on work release six days a week for up to 16 hours a day.

Edwards and other attorneys fought in court for a year before successfully getting the agreement unsealed in September 2009. More than 30 minor girls were identified as Epstein's victims in the pact.

PALM BEACH ESTATE
241 La Puerta Way, Palm Beach
Exquisitely designed Mediterranean Palm Beach walled estate. >> More Details
corcoran.com

TOWN OF PALM BEACH
1911 • CENTENNIAL • 2011
100 Click here for the latest events! 100

Find us on Facebook

Palm Beach Daily News
Like

2,413 people like **Palm Beach Daily News**.

Stanton

Jackie

Todd

Diana

Joe

Goharians

James

Karen

Facebook social plugin

MOST RECENT ALBUMS

The Palm Beach Opera

The Society of the Four Arts benefit

Pamella Roland

More »

Doe 1 and 2, who were 14 and 13, respectively, at the time of the incidents, received monetary settlements in civil cases. They are among more than two-dozen underage girls who filed lawsuits or settled claims against Epstein, alleging they were lured to his Palm Beach mansion to give him sexually charged massages and/or sex in exchange for money.

The motion filed Monday says the agreement is illegal because the government did not protect the “Congressionally mandated rights of victims before it entered this agreement.”

Alicia Valle, special counsel for the U.S. Attorney’s Office Southern District of Florida, said in an e-mail that the U.S. Attorney’s Office will respond in court filings.

“However, as we stated more than two years ago in July 2008 in our response to the plaintiffs’ then-emergency petition for enforcement of the Crime Victim Rights Act, the CVRA was not violated because no federal charges were ever filed in the Southern District of Florida,” Valle said. “Because the matter remains pending in court, it would be inappropriate at this time to provide additional comment on the merits of the current motion.”

The attorneys reference e-mails and letters from the federal office to Epstein’s lawyers acknowledging the government’s legal obligation to inform victims about the pact. The e-mails are redacted in the motion because they are under seal. The attorneys filed a separate motion Monday to unseal the correspondence.

“The reasonable inference from the evidence is that the U.S. Attorney’s Office wanted to keep the agreement a secret to avoid intense criticism that would surely ensued had the victims and the public learned that a billionaire sex offender with political connections had arranged to avoid federal prosecution for numerous felony sex offenses against minor girls,” the motion says. “As part of this pattern of deception, the U.S. Attorney’s Office discussed victim notification with the defendant sex offender and, after he raised objections, stopped making notifications.”

Epstein sought “a higher level of review” within the Department of Justice, the motion says. “A reasonable inference from the evidence is that Epstein used his significant political and social connections to lobby the Justice Department to avoid significant federal prosecution,” the motion states.

Share this article:

COMMENTS

Comments are closed

NEWS
Religion
Archives
ProfilePalmBeach
Madoff
Special Reports
Town Council
Feral Cats
Plaza
Reach 8
Health

SOCIETY
Insider
Social Calendar
FASHION
Fashion Calendar
OPINION
Local Voices
Letters to the Editor
Submit a Letter

WEATHER
SPORTS
Polo
WEF
BUSINESS
BLOGS
COLUMNISTS

ARTS
Arts Calendar
OBITUARIES
LIFESTYLES
Announcements
Food
Worth Avenue
Pets

ADVERTISE
SERVICES
Subscriber Services
Help & FAQ
Locations
Reprints
Staff
Privacy Policy
About Us
Visitor Agreement
Subscribe
REAL ESTATE

SPECIAL SECTIONS
Palm Beach Life
Home & Loggia
Automotive & Yacht
Showcase
Visitors' Guide
Season in Review
Hurricane Guide 2010
Chamber of Commerce
Newsletter

Copyright © Thu Apr 07 13:33:34 EDT 2011 All rights reserved. By using PalmBeachDailyNews.com, you accept the terms of our visitor agreement. Please read it.
Contact PalmBeachDailyNews.com | Privacy Policy | About our ads

ACAP ENABLED

Start or Renew Your Subscription

Search for Palm Beach Daily News

Follow @ShinySheet on Twitter!

Palm Beach Daily News ELECTRONIC EDITION Electronic Edition Now Available! Click Here To Subscribe

NEWS
Religion | Archives

E-mail this page Print this page Most popular

New Search Return to results Printer Friendly

About your archives purchase:
Your purchase of **20** articles expires on **04/07/2011 4:22 PM**.
You have viewed **3** articles and have **17** articles remaining.

Palm Beach Daily News (FL)

JUDGE RECEIVES EPSTEIN TAPE RULING PENDING

MICHELE DARGAN, Daily News Staff Writer

Published: May 5, 2010

NEW YORK -- A Manhattan federal judge Tuesday took into custody a tape-recorded conversation between veteran newspaper reporter George Rush and convicted sex offender Jeffrey Epstein.

But U.S. District Judge Lawrence M. McKenna reserved ruling on whether the recording will be released to attorneys representing young women who were sexually abused by Epstein as minors. McKenna didn't listen to the recording during the hearing. Fort Lauderdale attorney Brad Edwards and Utah attorney and law professor Paul Cassell are fighting to obtain the 22-minute tape on behalf of Epstein victim Jane Doe. She has filed one of a dozen pending civil cases in federal court in West Palm Beach against Epstein.

A status check is set for Thursday in those cases before U.S. District Judge Kenneth Marra.

Doe could have settled the lawsuit for $50,000 but is asking for $50 million in damages, Cassell said Tuesday. "Jane Doe was repeatedly sexually assaulted over a lengthy period of time by this wealthy and powerful man," Cassell said.

Epstein, 57, is currently under house arrest in his Palm Beach home after serving 13 months of an 18-month state sentence for soliciting a minor for prostitution and soliciting prostitution. Nearly two dozen young women have filed lawsuits against the billionaire money manager -- some already settled -- all alleging Epstein sexually abused them as minors at his El Brillo Way home.

Cites reporter's protected privilege

Rush, of the New York Daily News, was present in the courtroom, but did not have to testify Tuesday. Neither did Fort Lauderdale private investigator Michael Fisten, also in the courtroom. Working on behalf of Epstein victims, Fisten discovered the existence of the tape and had a conversation with Rush about its contents.

Representing Rush and the newspaper, Washington attorney Laura Handman and New York Daily News attorney Anne Carroll argued the tape should not be released under any circumstances, citing reporter's protected privilege. Rush told Epstein the conversation was "off the record" and has never published any portion of that conversation. But even if portions had been printed, the unpublished portions would still be protected, Handman said.

Handman cited cases where interviews were conducted in the presence of other people and privilege was not waived.

In addition, Handman argued that Rush should not have to testify in court.

The ability for reporter's privilege to be protected is crucial in culling sources and gathering information for news stories, Handman said. Reporting is all about give and take between the reporter and the source; that's what reporters have to do, Handman said.

"This is so critical to news gathering," Handman said. "Mr. Rush could find himself testifying in [many] cases just because he had the temerity to do some reporting on a very important story."

There is nothing helpful to Doe's case on the tape and "Jane Doe is not referred to once in that tape," Handman said.

Cassell argued that the tape is "critical in showing Epstein's lack of remorse."

Cassell described Epstein as a pitiless sexual abuser to Jane Doe and at least 30 other minor girls. Even though Jane Doe is not referred to by name on the tape, Epstein refers to his victims as "the girls" and makes disparaging remarks about them on the tape, Cassell said.

Tape played for others

Cassell said privilege does not apply because it was waived when Rush played the tape for three people and verbally divulged its contents to two others, including Fisten and Edwards, who also represents two other victims. But even if there is "qualified privilege," Cassell says, it is outweighed by Doe's inability to obtain the information anywhere else and the jury's need to hear Epstein's own words about his lack of remorse.

Since Epstein has exercised his Fifth Amendment right during questioning by victims' attorneys, the jury

Classifieds Real Estate
Advertise Automotive

THE INSIDER BLOG
by Shannon Donnelly

will have no other way to hear Epstein's words in his own voice, Cassell said.
Deadline for discovery in the Doe case is May 31, with the trial set for July 14.
-- mdargan@pbdailynews.com

Copyright (c) 2010 Palm Beach Daily News

Refinance Rates at 2.65%
$160,000 Mortgage for $659/mo. No SSN req
LendGo.com/Mortgage

Do NOT Buy Car Insurance
We found out how drivers can get
www.News7BreakingNews.com

Man "Cheats" Credit Score
He Added 126 Points To His Credit Score
www.CreditRepairFromHome.com

$79/Hr Job - 262 Openings
Make $79/hr Working From Home. As seen on
www.workfromhomeguide.net/jobs

Ads by Yahoc

By using this service you accept the terms of our Visitor Agreement.
Copyright 2007 Palm Beach Daily News. All rights reserved.
The Palm Beach Daily News
Privacy Policy | About this site | Write to us

ACAP ENABLED
COXMEDIA GROUP

Start or Renew Your Subscription

Search for Palm Beach Daily News

Follow @ShinySheet on Twitter!

JUST A CLICK AWAY! Dining Directory PalmBeachDailyNews.com

Classifieds | Real Estate | Advertise | Automotive

updates in your e-mail

Sign up to receive our e-mail newsletters here.

SUBSCRIBE

Palm Beach Daily News NEWSLETTERS

About your archives purchase:
Your purchase of **20** articles expires on **04/07/2011 4:22 PM**.
You have viewed **2** articles and have **18** articles remaining.

Palm Beach Daily News (FL)

LAWYER: EPSTEIN MADE ADMISSIONS ON TAPE

MICHELE DARGAN, Daily News Staff Writer

Published: April 29, 2010

A tape recorded interview between a reporter and convicted sex offender Jeffrey Epstein contains "damning admissions by Epstein," which includes Epstein saying he had come "close to crossing a line" concerning sex with underage girls.

Those and other revelations about the 22-minute interview by New York Daily News reporter George Rush with Epstein are contained in a 24-page court filing by attorneys Brad Edwards and Paul Cassell on behalf of Jane Doe. Edwards represents Doe and two other Epstein victims. Edwards and Cassell are fighting to obtain the tape to further their case of sexual abuse by Epstein when Doe was a minor.

Epstein, 57, was released from jail in July after serving 13 months of an 18-month state sentence for soliciting a minor for prostitution and soliciting prostitution. Nearly two dozen young women have filed lawsuits against Epstein -- some already settled -- all alleging Epstein sexually abused them as minors at his El Brillo Way home, where he is now serving house arrest.

The New York Daily News is seeking to keep the tape confidential, citing reporter's protected privilege.

In response, Cassell says the newspaper waived its protected privilege when Rush played the recording for three people and described its contents to two others, including Edwards.

In addition, Cassell writes that privilege can't be applied in this situation because it doesn't involve an issue related to a confidential source. The person on the tape is Epstein.

Even if there is "qualified privilege," Cassell maintains it is outweighed by Doe's inability to obtain the information anywhere else and her "compelling need to obtain Jeffrey Epstein's own words about his sexual abuse and lack of remorse."

When reached by phone Wednesday, Anne Carroll, attorney for the New York Daily News, said she will answer Cassell in a court filing. Both the New York Daily News and Doe have asked a federal court judge in Manhattan to listen to the tape in chambers to help determine whether privilege applies.

Epstein and others who helped him procure minor girls for massages and sex acts have taken the Fifth Amendment in their depositions, stymieing Doe and the other victims suing Epstein, the documents say.

Michael Fisten, an investigator working for Doe, discovered the existence of the tape in fall 2009. An author who had listened to the tape told Fisten that Rush had a tape recording of Epstein "discussing the sexual abuse of minor girls."

According to a sworn affidavit by Fisten, he called Rush, who confirmed he interviewed Epstein and made a tape. According to Fisten, Rush told him that he compiled negative information from Epstein about his exploits with underage girls and how he eluded the justice system. But Fisten said that Rush told him that his publisher, who knows Epstein, killed it after receiving a call from Epstein.

Fisten said Rush told him, among other things, that the following information was contained on the tape:

That Epstein said he went to jail in Florida for no reason and if the sexual abuse of minors had happened in New York, he would have only received a $200 fine.

That L.M., one of Edwards' clients who sued Epstein for sexual abuse as a minor, came to him as a prostitute and a drug user (meaning she came to him for sex, rather than him pursuing her).

That all the girls suing him are only trying to get a meal ticket.

That the only thing he might have done wrong was to maybe cross the line a little too closely.

In a sworn deposition, Edwards states that Rush disclosed much of the information contained on the tape to him in a conversation.

Edwards said in his statement that the Rush interview is "unique and not otherwise obtainable from other witnesses because it can be used to prove perjury -- a federal crime.

Edwards said Epstein testified in a deposition that he did not recognize the name George Rush from the New York Daily News "despite the fact that he gave a personal interview that we all now know to have been tape recorded."

-- mdargan@pbdailynews.com

Epstein
Has 'come close to crossing a line.'

Copyright (c) 2010 Palm Beach Daily News

Refinance Rates at 2.65%
$160,000 Mortgage for $659/mo. No SSN req
LendGo.com/Mortgage

LEAKED: $25 Car Insurance
Your Auto Insurer hates this.
News7BreakingNews.com

Man "Cheats" Credit Score
He Added 126 Points To His Credit Score
www.CreditRepairFromHome.com

$79/Hr Job - 262 Openings
Make $79/hr Working From Home. As seen on
www.workfromhomeguide.net/jobs

Ads by Yahoc

By using this service you accept the terms of our Visitor Agreement.
Copyright 2007 Palm Beach Daily News. All rights reserved.
The Palm Beach Daily News
Privacy Policy | About this site | Write to us

ACAP ENABLED
COXMEDIA GROUP

From: Darren Indyke <[redacted]>
To: Jackie Perczek [redacted]
Date: 4/7/2011 1:27 PM
Subject: Privileged and Confidential
Attachments: Edwards Articles -1.doc; Edwards Articles - 12.pdf; Edwards Articles - 10.pdf; Edwards Articles - 9.doc; Edwards Articles - 8.doc; Edwards Articles - 7.doc; Edwards Articles - 6.doc; Edwards Articles - 5.doc; Edwards Articles - 3.doc; Edwards Articles - 2.doc; Edwards Article - 13.pdf; Part.012

See additional articles.

Darren K. Indyke
Darren K. Indyke, PLLC
301 East 66th Street, 10B
New York, New York 10065
[redacted]

The Palm Beach Post
NEWS

AdChoices
LEARN MORE ›
LOCATE ›
NEXT GENERATION PENNZOIL®
CLEANS OUT UP TO 40% OF SLUDGE IN THE 1ST OIL CHANGE.*
*Based on a severe sludge clean-up test using SAE 5W-30.
Not just oil, Pennzoil.®

BREAKING NEWS: Gov. Scott OKs last-minute bailout for courts, averting two-week furloughs Click to read story

Palm Beach sex offender's secret plea deal: Possible co-conspirators not charged, presses victims to settle civil suits

By SUSAN SPENCER-WENDEL

Palm Beach Post Staff Writer

Friday, September 18, 2009

WEST PALM BEACH — Billionaire financier sex offender Jeffrey Epstein's secret non-prosecution agreement he struck with federal prosecutors was unsealed Friday, offering the first public look at the deal Epstein's high-powered legal counsel brokered on his behalf.

According the agreement, the Federal Bureau of Investigation and the U.S. Attorney's Office investigated Epstein for various federal crimes, including prostitution, some punishable by a minimum of 10 years up to life in prison

But federal prosecutors backed down and agreed to recall grand jury subpoeanas, if Epstein pleaded guilty to prostitution-related felonies in state court, which he ultimately did. He received an 18-month jail sentence, of which he served 13.

A former federal prosecutor of 15 years, Mark Johnson of Stuart, said the disparity in the potential sentences was unusual.

The United States Attorney's Office also agreed not to charge any of Epstein's possible co-conspirators - Sarah Kellen, Adriana Ross, Lesley Groff and Nadia Marcinkova.

The agreement was negotiated in part by New York heavyweight criminal defense attorney Gerald Lefcourt.

On its first draft in September 2007, it required that Epstein pay an attorney - tapped by the U.S. Attorney's Office and approved by Epstein - to represent some of the victims in civil suits they had filed against Epstein. That attorney is prominent Miami lawyer Bob Josefsberg.

Former prosecutor Johnson said he has never seen a provision like that before.

But an addendum to the agreement signed the following month struck Epstein's duty to pay Josefsberg if he and the victims did not accept a settlement and instead pursued litigation.

The agreement, signed by Assistant U.S. Attorney Maria Villafana, does not expressly state whether any victims were contacted or consulted before the deal was made.

Attorney Brad Edwards of Fort Lauderdale, who represents three of the young women, believes that none of the between 30 and 40 woman identified as victims in the federal investigation were told of the deal. Edwards said his clients were still receiving letters in the mail months afterwards saying the U.S. Attorney's Office assuring them Epstein would be prosecuted.

"Never consulting the victims is probably the most outrageous aspect of it..." Edwards said. "It taught them that someone with money can buy his way out of anything. It's outrageous and embarrassing for United States Attorney's Office and the State Attorneys Office."

Epstein now faces many civil lawsuits filed by the women, who are represented by a variety attorneys. In many, the facts alleged are the same: that Epstein had a predilection for teenage girls, identified poor, vulnerable ones and lured them to his home via other young women. The teens describe ascending a staircase lined with nude photographs of young girls and to the spa room where Epstein would appear in a small towel.

Former Circuit Judge Bill Berger, who represents one of the victims, and The Palm Beach Post sought the unsealing of the agreement. Berger refers to it as a "sweetheart deal."

"Why was it so important for the government to make this deal?" Berger asked rhetorically. "We have not yet had honest explanation by any public official as to why it was made .. and why the victim's were sold down the river."

Former federal prosecutor Ryon McCabe described the agreement as "very unorthodox." Such agreements, he said, are usually reserved for corporations, not individuals.

Obama Launches Mortgage Relief Plan
If you owe less than $729,000 on your mortgage, you probably qualify for the President's Making Home Affordable Program. With rates lower than they've ever been, there has never been a better time to refinance. If you are a homeowner and you haven't looked into refinancing recently, you may be surprised at how much you can save.
Select Your Age: 33
Calculate New House Payment

COLUMNISTS AND BLOGGERS

FRANK CERABINO
Read Frank's latest columns and follow him on Twitter. Read more

HOT CELEBRITY NEWS
Get the latest on South Florida celebrities, billionaires, politicos, more. Page2Live

GEORGE BENNETT
Read Post politics columnist George Bennett's latest articles. Read more

MOST POPULAR HEADLINES COMMENTS
Fatal shooting in Delray Beach draws crowd of 100 onlookers
Lake Worth mayor says The Cottage complaints use 'gay card' against city manager
West Palm Beach mayor: Firefighter layoffs likely
Narcy Novack charged in 2009 Fort Lauderdale killing of her mother-in-law
Boynton Police warn of new twist on ATM identity fraud

FOLLOW THE POST ON TWITTER
SIGN UP FOR MOBILE TEXT ALERTS

The Palm Beach Post on Facebook
Like
16,985 people like The Palm Beach Post.

"It's very, very rare. I've never seen or heard of the procedure that was set up here." said McCabe, who has no involvement in any Epstein litigation and is now a securities litigation attorney.

"He's essentially avoiding federal prosecution because he can afford to pay that many lawyers to help those victims review their cases.... If a person has no money he couldn't be able to strike a deal like this and avoid federal prosecution."

The back-room deal with federal prosecutors all the more interesting in light of the legal heavyweights who have worked for Epstein, including Harvard professor Alan Dershowitz and Kenneth Starr of Clinton impeachment fame. Lefcourt is a past president of the National Association of Criminal Defense Lawyers.

Epstein's local defense attorney, Jack Goldberger, issued a statement Friday saying he had fought the release of the sealed agreement to protect the third parties named there. "Mr. Epstein has fully abided by all of its terms and conditions. He is looking forward to putting this difficult period in his life behind him. He is continuing his long standing history of science philanthropy..."

Epstein ended up avoiding federal charges, and pleaded guilty in state court to felony solicitation of prostitution and procuring a person under the age of 18 for prostitution. In July 2008, he was sentenced to 18 months in jail, and later allowed out up to six days a week on work release.

Epstein left the jail in late July 2009 after serving not quite 13 months of the sentence, having earned gain time for good behavior.

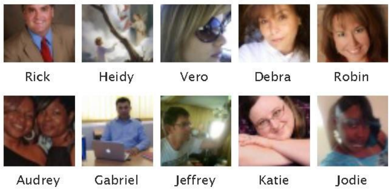
Palm Beach Police began investigating the "international moneyman of mystery," as the New York magazine dubbed him, after they received a complaint from a relative of a 14-year-old girl who had given Epstein a naked massage at his home on the Intracoastal Waterway.

Police sought and found in poor neighborhoods a variety of tall, thin, model-like young women, who told stories of begin recruiting, then going to Epstein's home and massaging and stimulating him. They walked away with between $200 and $1,000.

The investigation triggered tensions between police and prosecutors, with then-Chief Michael Reiter saying in a May 2006 letter to then-State Attorney Barry Krischer that the chief prosecutor should disqualify himself.

"I continue to find your office's treatment of these cases highly unusual," Reiter wrote. He then asked for and got the federal investigation that ended in the sealed deal.

"The Jeffrey Epstein matter was an experience of what a many-million-dollar defense can accomplish," Reiter told the Palm Beach Daily News upon his retirement.



Recent Activity

Login You need to be logged into Facebook to see your friends' activity

Cerabino: Florida House GOP's 'uterus' ban: A free-speech battle is born
1,518 people shared this.

Foreclosure crisis: Fed-up judges crack down disorder in the courts
207 people shared this.

Facebook social plugin

POSTPIX » Latest news photos

IMAGES OF WAR in Iraq and Afghanistan

Massive earthquake and tsunami devastate Japan

Jeri Muoio Sworn in as Mayor

Severe weather in Central, South Florida

Do Your Feet Hurt?
SELBY SHOES
561-969-9369
We'Ll Put Wow! Into Your Windows
IN THE SHADE INC
772-223-1212
Free Hearing Test!
BELTONE
561-948-3049
Gourmet Meat in Delray
MARIO'S MEATS
561-499-7019
Do Your Feet Hurt?
SELBY SHOES
561-969-9369

Historic Archive
(1897 - 1988)
Search historic editions of *The Palm Beach Post*, *Palm Beach Daily News*, *Miami News* and more. It's free!

SEARCH
Get this search box for your site.

UTOG: Hottest Oil Stock
Fracking tech to unlock vast treasure of 7,000-acre
www.AmericanEnergyReport.com

"Weird Fruit Burns Fat"
Reporter Drops 32 Pounds in 28 Days with This Strange
www.5NewsTV.com

Mortgage Rates Hit 2.99%
If you owe under $729k you probably qualify for Obama's
www.LowerMyBills.com

Ads by Yahoo!

Stay on top of the storms with our new interactive live radar – the radar you control

TELL US WHAT YOU THINK OR SEND US A TIP

Please use this box to submit general site feedback, technical problems or news tips. Thank you.

Your e-mail address (will not be shared)

SEND NOW

SECTIONS

Local News
Money
Sports
Weather
Opinion
Traffic
Special reports
Post on Politics
Entertainment pbpulse.com
Treasure Coast Talk
Blogs
Twitter
Obituaries
Living
Photos - PostPix
Videos

MARKETPLACE

Real Estate Florida Home
Local Businesses
Advertise in The Post
Advertise on PalmBeachPost.com

SERVICES

Customer Care
Home Delivery
Local business directory

ARCHIVES

Front page PDFs
Historic Archives

AFFILIATED SITES

GalleryPalmBeach.com
HistoricPalmBeach.com
PalmBeachDailyNews.com
pbgametime.com
pbpulse.com
Page2Live.com
DoGood - Local non-profits

ON THE GO

E-mail Newsletters
RSS Feeds
Mobile Services
Reprints/licensing

Copyright © 2011 The Palm Beach Post. All rights reserved. By using PalmBeachPost.com, you accept the terms of our visitor agreement. Please read it.
Contact PalmBeachPost.com | Privacy Policy | About our ads

ACAP ENABLED

The Palm Beach Post
NEWS

BRAMAN MOTORCARS PALM BEACH
EXPERIENCE THE BRAMAN DIFFERENCE
Porsche
Mini
BMW
Audi

Local News *Greater Palm Beaches and Treasure Coast*

SEARCH
○ Site ○ Web Web Search by YAHOO!

Judge: Palm Beach sex offender Jeffrey Epstein agreement to remain sealed

By MICHELE DARGAN

Palm Beach Daily News Staff Writer

Tuesday, February 17, 2009

A federal judge has ruled that a non-prosecution document under which the government agreed not to pursue federal charges against sex offender Jeffrey Epstein will remain under seal - at least for now.

The U.S. Attorney's Office and Epstein's lawyers reached the agreement before Epstein pleaded guilty to state felony charges, and the document is under seal in Epstein's state criminal file.

Representing two of Epstein's victims, attorney Brad Edwards asked to have the document unsealed as part of his federal lawsuit against the Manhattan money manager. Although Edwards and his victims have seen the agreement, Edwards says in his pleadings that the government has "inaccurately described the agreement ... creating a false impression that the agreement protects the victims."

U.S. District Judge Kenneth Marra ruled that the claims, even if true, haven't damaged Edwards' case.

"If and when such alleged mischaracterizations become relevant to an issue to be decided by the court, the parties will be given the opportunity to advance their positions and the court will resolve the issue," he wrote. "If disclosure of the agreement will be required for the court to resolve this issue, appropriate disclosure will be ordered."

Seeking to keep the agreement sealed, Assistant U.S. Attorney Dexter Lee argued that the agreement is not part of any case before Marra.

"The non-prosecution agreement has never been filed under seal in federal court," he wrote.

He also denied that the agreement has been inaccurately described.

Marra sided with Lee on the argument that the agreement was not filed in federal court "under seal or otherwise."

On Aug. 14, Marra ruled that the non-prosecution agreement would be unsealed for Edwards and any of the victims who want to see it. But the ruling bars Edwards and anyone else who sees the document from disclosing the terms to anyone else.

In his motion to unseal, Edwards said he wants to be able to discuss the terms of the agreement with other victims and their attorneys as well as with other victims' rights groups such as the National Alliance of Victims' Rights Attorneys.

The desire to discuss the agreement with third parties is not justification for unsealing the document, Marra ruled.

"If a specific tangible need arises in a civil case ... relief should be sought in that case," he wrote.

Epstein, 56, is serving 18 months in jail for soliciting prostitution and procuring a minor for prostitution.

Under the agreement, federal prosecutors will defer their decision on whether to prosecute Epstein on federal charges until 90 days after Epstein completes all requirements of his state sentence.

If he abides by all court conditions and restrictions, the federal case would be dropped.

In addition to the state criminal case, there are nine federal and seven state lawsuits pending against Epstein.

All contain similar allegations: The Manhattan money manager, through his employees and assistants, brought minor girls to his Palm Beach home at 358 El Brillo Way for erotic massages and sometimes sex.

ACCOUNTS RECEIVABLE SPECIALIST
Busy Physical Therapy office seeks FULL TIME AR Collection person to handle patient accounts.
Click to Apply!
EXPERIENCE PREFERRED

COLUMNISTS AND BLOGGERS

FRANK CERABINO
Read Frank's latest columns and follow him on Twitter. Read more

HOT CELEBRITY NEWS
Get the latest on South Florida celebrities, billionaires, politicos, more. Page2Live

GEORGE BENNETT
Read Post politics columnist George Bennett's latest articles. Read more

MOST POPULAR HEADLINES COMMENTS

Fatal shooting in Delray Beach draws crowd of 100 onlookers
Lake Worth mayor says The Cottage complaints use 'gay card' against city manager
Narcy Novack charged in 2009 Fort Lauderdale killing of her mother-in-law
West Palm Beach mayor: Firefighter layoffs likely
Boynton Police warn of new twist on ATM identity fraud

FOLLOW THE POST ON TWITTER
SIGN UP FOR MOBILE TEXT ALERTS

The Palm Beach Post on Facebook
Like

16,987 people like The Palm Beach Post.

Jodie

Heidy

Bianca

Robin

Audrey

Katie

Gabriel

Vero

Yelena

Debra

Recent Activity

Login You need to be logged into Facebook to see your friends' activity

Cerabino: Florida House GOP's 'uterus' ban: A free-speech battle is born
1,524 people shared this.

Foreclosure crisis: Fed-up judges crack down disorder in the courts
207 people shared this.

Facebook social plugin

POSTPIX » Latest news photos

IMAGES OF WAR in Iraq and Afghanistan

Massive earthquake and tsunami devastate Japan

Jeri Muoio Sworn in as Mayor

Severe weather in Central, South Florida

FEATURED MOTORCYCLES

1997 Harley Davidson Road King 941-769-9000
HD Road King '97, 48K...

View All Featured Motorcycles

Historic Archive
(1897 - 1988)

Search historic editions of *The Palm Beach Post*, *Palm Beach Daily News*, *Miami News* and more. It's free!

SEARCH

Get this search box for your site.

"Weird Fruit Burns Fat"
Reporter Drops 32 Pounds in 28 Days with This Strange
www.5NewsTV.com

US Oil Best Kept Secret?
Not for long. Similar stocks trade at $30-$60.
www.AmericanEnergyReport.com

BREAKING NEWS:$25 Car Ins
Do Not Buy Car Insurance until you see this shocking
www.News7BreakingNews.com

Ads by Yahoo!

Stay on top of the storms with our new interactive live radar – the radar you control

LexisNexis®

7 of 11 DOCUMENTS

Copyright 2009 ProQuest Information and Learning
All Rights Reserved
ProQuest SuperText
Copyright 2009 Palm Beach Post
Palm Beach Daily News

June 25, 2009 Thursday
Final Edition

SECTION: A SECTION; Pg. A.1

LENGTH: 557 words

HEADLINE: JUDGE TO RULE ON SEALED PLEA-DEAL PAPERS TODAY

BYLINE: MICHELE DARGAN, MICHELE DARGAN, Daily News Staff Writer

BODY:

A circuit judge will decide today whether the public will be privy to the federal government's non-prosecution deal with Jeffrey Epstein, which was sealed when the convicted sex offender pleaded guilty in June 2008 to two felony counts.

Epstein, of Palm Beach, will be released from the Palm Beach County Stockade July 22, after serving less than 13 months of his 18- month sentence for procuring a minor for prostitution and solicitation of prostitution.

Teri Barbera, spokeswoman for the Palm Beach County Sheriff's Office, confirmed his release date Tuesday.

Epstein's projected release date had been Sept. 24, but gain time -- which includes his participation in a work-release program -- moves the date up to July 22, Barbera said.

Epstein, 56, has been in the work-release program since Oct. 10, in which he is allowed out of the stockade six days a week, from 10 a.m. to 10 p.m., to go to his West Palm Beach office, the Florida Science Foundation, monitored by an ankle bracelet and accompanied by a deputy.

As part of Epstein's state plea agreement, the U.S. Attorney's Office agreed not to prosecute Epstein on federal charges as long as he fulfills all requirements of his sentence and probation. The federal non-prosecution agreement has been under seal in state court.

Epstein's attorney Jack Goldberger filed court papers asking that the documents stay sealed for the following reasons: "to prevent a serious imminent threat to the fair, impartial and orderly administration of justice; to protect a compelling government interest; to avoid substantial injury to innocent third parties and to avoid substantial injury to a party by disclosure of matters protected by a common law and privacy right, not generally inherent in these specific type of proceedings, sought to be closed."

Fort Lauderdale-based attorney Brad Edwards represents three Epstein victims and has asked Circuit Judge Jeffrey Colbath to unseal the federal agreement to the public. An attorney for The Palm Beach Post also has asked that the records be unsealed.

Edwards and his clients have seen the agreement after a federal judge ruled that they are allowed to see it. But that ruling bars Edwards and anyone else who sees the document from disclosing the terms to anyone else.

Edwards said he wants to use that document "in the deposition of various material witnesses" relative to his cases.

Radaronline.com has reported that Epstein has "secretly been helping the feds unravel a Ponzi scheme" related to the June 2008 indictment of two former managers of Bear Stearns Mortgage Investment Fund.

Epstein's rep, Howard Rubenstein, confirmed last year that Epstein is "Major Investor No. 1" in the indictment, which says he lost about $57 million.

Goldberger could not be reached for comment.

The Manhattan money manager has been incarcerated since June 30, when he pleaded guilty to the two felony counts. As part of the plea agreement, Epstein must serve one year of house arrest after his release and register as a life-long sex offender.

In addition to the criminal case, there are more than a dozen civil lawsuits -- both state and federal -- pending against Epstein. All contain similar allegations: Epstein, through his employees and assistants, brought minor girls to his Palm Beach home on El Brillo Way for erotic massages and sometimes sex.

-- mdargan@pbdailynews.com

GRAPHIC: Caption: Epstein To be released from jail July 22.

LOAD-DATE: September 1, 2010

NEW YORK POST
Last Updated:Wed., Apr. 6, 2011, 01:45pm

BERGDORF GOODMAN
LANVIN
SHOP NOW

Derek Jeter Carmelo Anthony Jay-Z Donald Trump WEEKLY AD PHOTOS VIDEOS BLOGS

News

Phrase-y Charlie
He wants to leave his mark -- on your wallet. Warlock wannabe...

Page Six

- Padma jabs
- John's ex wastes no time
- Lots of attention

Sports

Poison pen
One, two, three flush. Asked to protect a gem by CC Sabathia,...

Entertainment

Kitchen's hot!
On a recent Saturday evening, a small army of servers outfitted...

NYC Local Business Opinion Teams High Schools Scores TV Movies Events Travel

Home Cindy Adams Celeb Photos PopWrap Fashion Delonas Cartoon Page Six Magazine

Story
Comment

Heiress quizzed in sex suits

Last Updated: 3:35 AM, October 12, 2009
Posted: 12:55 AM, October 12, 2009
Comments: 14
Like Be the first of your friends to like this.
0

| More Print

Ghislaine Maxwell, the British brunette whose father once owned the Daily News, has been slapped with a subpoena in suits brought by 24 underage girls against her old friend, billionaire **Jeffrey Epstein**.

Maxwell -- whose press-lord father, Robert Maxwell, died in 1991 after falling into the Atlantic off his yacht, the Lady Ghislaine -- was served with a subpoena on Sept. 22 at 6:45 p.m. as she was leaving the Clinton Global Initiatives Conference at the Sheraton Hotel.

Florida lawyer **Brad Edwards**, who represents three of the "Jane Does" who are suing Epstein, told Page Six that Maxwell would be questioned over her knowledge of how Epstein procured many of the girls.

Epstein is accused in the civil complaints of luring underage girls to his mansion in Palm Beach to give him massages, during which he allegedly engaged them in sexual activity and paid them hundreds of dollars each. A grand jury indicted Epstein on a charge of felony solicitation of prostitution. Epstein, who pleaded guilty and did 12 months in prison, was deposed last week in the offices of his lawyer, **Jack Goldberger**.

Goldberger wouldn't comment, but a friend of Epstein said, "These [people bringing the complaint] are the lowest of the lows in terms of ambulance-chasing lawyers." The trials are scheduled to start in February.

Nadia Marcinkova, who has been described as Epstein's lesbian sex slave and who visited him behind bars 67 times, has also been served with a subpoena.

Epstein's brother, **Mark Epstein**, who has a real-estate holding company in New York, has already been deposed about a building he owns, 301 E. 66th St. "Jeffrey rents several apartments there where he keeps his girls, alleged models for the MC2 agency he owns," Edwards said. "But Mark acts like he doesn't even know his brother. He was extremely angry and rude and cursed me out."

Emily Smith
with Ian Mohr, Stephanie Smith and Tara Palmeri

Cindy Adams

- **Bravo stars snicker at Padma**
- **John's ex wastes no time**
- **Lots of attention**
- **Italian approach**
- **'Kennedys' hunt**
- **Mariah Carey shows off pregnant belly**
- **Private politics**
- **Quick to buy**
- **Madge rips 'false' stories**
- **Easy charmer**
- **Amar'e is selling his Miami duplex**
- **Parking struggle**
- **We hear**
- **Sightings**

First Look: D-Trix judges 'ABDC'
If you know him from "So You Think You Can Dance" season three or "America's Best Dance...

Kelly Clarkson's greatest hits sound greater than ever
The Clarkson colors run deep here at PopWrap, so

View Singles in New York
VIEW PHOTOS OF: Men NEAR ZIP: VIEW PHOTOS

seeking1178, age 36
Seeking: women 24-35
Interests: Camping, Cooking, Dining out, Wine tasting
See More Like Him

aaron4255, age 27
Seeking: women 23-33
Interests: Dining out, Exploring new areas, Music and concerts
See More Like Him

dar5508, age 31
Seeking: women 23-35
Interests: Nightclubs/ Dancing, Playing sports, Travel / Sightseeing
See More Like Him

match.com

Sponsored Links Buy a Link Here

Scholarships for Moms
Grant Funding May Be Available to Those Who Qualify!
SeeCollegeDegrees.com/Grant

Mom Shares $3 Wrinkle Tip
Dermatologists HATE this 1 simple tip that erases wrinkles fast
WeeklyInform.com

SHOCKING: 52 Samsung HDTV for $74.07
TODAY ONLY: Auction site to clear out 1,000 52 Samsung HDTVs for $74.07!
Consumer-Weekly.net/SmartShopping

Like Be the first of your friends to like this.

0

| More Print

Comments

Add a comment...

Comment

Facebook social plugin

NYPOST Comments (10)

Jimsco16 Report
03/28/2011 9:26 AM
No Chance I'm going to believe any of this. "Jane Does" seems to be the operative phrase in all of this. Do a google search for "Ghislaine Maxwell Vanity Fair" you'll find a relatively different, much more founded story there.

Pedro Gonzalez Report
10/12/2009 8:59 PM
...hatinTHEHATERS....You have just shown what kind of hypocrites use the Race Card......

shodan Report
10/12/2009 7:43 PM
hatinTheHaters: Wow, blaming the underages girls he molested? Word to the wise for the other posters here. Keep your kids away from "hatinTheHaters".

Kramden Report
10/12/2009 3:10 PM
Jeffrey Cumstein?

Yankee Doodle Dandy Report
10/12/2009 11:50 AM
The New York Post NEVER misses an opportunity to tweak the nose of its cross-town rival, The New York Daily News!
Hey, if you CAN'T bash 'em, TRASH 'em!!!

jcmrtn Report
10/12/2009 11:15 AM
But Hollywood types say raping and drugging 13 year olds isn't really a crime anymore. Jeffrey and Roman should be released and given an apology. How dare we have laws to protect children. Don't they realize these men are FAMOUS?

palebluehalo Report
10/12/2009 7:05 AM
c'mon, who hasn't banged a few underage models?

Pedro Gonzalez Report
10/12/2009 5:07 AM
...Ghislaine Maxwell's mistake was to be at the Clinton Global

Kelly's long-teased appearance on today's...

This 'When Harry Met Sally' sequel sucks
Although I haven't seen it in quite a number of years, I suspect that "When Harry Met...

'30' to 'Rock' no more
Alec Baldwin is going from "Rock" to "Rock" as the "30 Rock" star preps for his role in an...

Back to Bayside
...

Good grief, Jennifer Garner is such a dork
Jennifer Garner may have a body, life and closet to kill for, but every time she does...

Watch two minutes from 'Workaholics'
Comedy Central is very hit and miss with their original programming -- for every "South...

More in **PopWrap**

Sponsored Links

53yr Old Mom Looks 30!
We Expose the $4 Trick to ERASE Wrinkles. Her Results will Shock You!
www.vanityreports.com/...

Obama Urges Homeowners to Refinance
If you owe under $729k, you probably qualify for...
www.SeeRefinanceRates....

Hot Stock Pick - GTSO
New Rare Earth Exporter in Mongolia Prepares to Ship New Ore Next Mont
www.RareEarthExporters...

Buy a link here

More in **Videos**

News | Gossip | Sports | Weird News | Lifestyle

NEW YORK POST "The Best iPad App" ★★★★★ Available on the App Store

FOLLOW THE NEW YORK POST ON twitter

Initiative Conference, a tax dodge for the rich and a front for the sexual deviants. The place was no doubt under surveillance by the IRS and the FBI Pedophile unit.

1 2 **Next» Last»**

See All Comments

You must be logged in to leave comments. **Login** | **Register**

POST SECTIONS

News: Business | Opinion | Columnists | Local News
Sports: Columnists | Scores | Fantasy
PageSix: Celeb Photos | POPWRAP | Page Six Magazine
Entertainment: TV | Movies | Music
Classifieds: Rentals | Jobs | Cars | Real Estate | NY Apartments
Multimedia: Photos | Video
Specials: Contests/Sweepstakes

CUSTOMER CARE

Contact Us
FAQ
Today's Paper
Archives
Covers
Back Issues
Reprints
Sitemap
Help

SUBSCRIBE

Home Delivery
iPad
Daily Newsletter
iPhone
E-Edition
Mobile
RSS

ADVERTISING/PARTNERS

Media Kit
Parade Magazine
Coupons

NEW YORK POST

Can't find what you're looking for? Try searching:

NEW YORK POST is a registered trademark of NYP Holdings, Inc.
NYPOST.COM, NYPOSTONLINE.COM, and NEWYORKPOST.COM are trademarks of NYP Holdings, Inc.
Copyright 2011 NYP Holdings, Inc. All rights reserved. Privacy | Terms of Use

9 of 13 DOCUMENTS

Copyright 2009 Sun-Sentinel Company
All Rights Reserved

Sun-Sentinel (Fort Lauderdale, Florida)

June 15, 2009 Monday
Palm Beach Edition

SECTION: LOCAL; Pg. 3B

LENGTH: 348 words

HEADLINE: HEARING SET TO CONSIDER SECRECY OF PLEA BARGAIN

BYLINE: Susan Spencer-Wendell The Palm Beach Post

BODY:

A Palm Beach Circuit Court judge will not immediately unseal a deal that wealthy Palm Beach money manager Jeffrey Epstein made with federal prosecutors to avoid charges.

Circuit Judge Jeff Colbath acknowledged, though, at a hearing last week that Epstein's deal was not sealed in accordance with state and local court rules.

"I don't see where any of the procedures were ever followed to begin with," Colbath said.

Colbath also set a full hearing on the matter for June 25.

Attorneys for young women now suing Epstein, together with The Palm Beach Post, are asking Colbath to unseal the deal that Epstein made with federal prosecutors.

"It's a secret agreement, a secret sweetheart agreement," said former Circuit Judge Bill Berger, who represents some of the women. "Everybody was in on this deal except the victims and the public. The public should be outraged it has gone as far as it has."

Brad Edwards, a second attorney representing the women, has seen the sealed deal after a federal judge allowed him and his clients to view it, but would not discuss its contents.

Edwards would say only that the women were "outraged" that it had been negotiated behind their backs.

A reporter asked Edwards whether he thought Epstein received special treatment by federal prosecutors.

"Are you kidding? It's transparent. Certainly, no one else gets treated like that," Edwards said.

Epstein, 56, a reported money manager of billionaires, is serving an 18-month sentence in the Palm Beach County Stockade after pleading guilty almost a year ago in state court to felony solicitation of prostitution and procuring teenagers for prostitution. Epstein is allowed out, though, each day from 7 a.m. to 11 p.m., a Sheriff's Office spokesman said.

Displeased with the way the State Attorney's Office handled the case, Palm Beach police forwarded information to the FBI.

INFORMATIONAL BOX:

Young women have sued

Money manager Jeffrey Epstein made a deal and is serving an 18-month sentence in jail. Attorneys for young women suing Epstein are asking a judge to unseal the deal that Epstein made with federal prosecutors.

NOTES: < Informational box at end of text. {TOPIC} Prostitution solicitation case

LOAD-DATE: June 15, 2009

3 of 3 DOCUMENTS

Copyright 2009 ProQuest Information and Learning
All Rights Reserved
ProQuest SuperText
Copyright 2009 Palm Beach Post
Palm Beach Daily News

June 11, 2009 Thursday
Final Edition

SECTION: A SECTION; Pg. A.1

LENGTH: 561 words

HEADLINE: EPSTEIN PLEA DEAL TO REMAIN SEALED FOR NOW

BYLINE: DAVID ROGERS, DAVID ROGERS, Daily News Staff Writer

BODY:

The plea deal that part-time Palm Beacher Jeffrey Epstein agreed to several months ago to avoid federal charges will remain sealed for the time being.

Lawyers for The Palm Beach Post and a woman who claimed Epstein solicited and procured her for sex at his El Brillo Way home while she was underage asked Palm Beach County Circuit Judge Jeffrey Colbath Wednesday morning to unseal the plea documents immediately.

Instead, Colbath decided to leave the documents sealed and give Jack Goldberger, the attorney representing Epstein, until 1 p.m. Friday to file papers showing why the records should remain out of public view.

Colbath agreed to let the Post and "E.W." have standing in the case and set a hearing on whether the documents should be unsealed for 1:30 p.m. June 25.

Epstein agreed in September 2007 to plead to state charges to avoid federal prosecution, Goldberger told County Judge Deborah Pucillo at Epstein's plea conference last year.

The sealing of the records in question was said to be a "significant inducement" for Epstein, who is serving 18 months in the Palm Beach County Stockade -- with daytime release -- and is facing several civil lawsuits in state and federal courts by more than a dozen alleged victims.

Colbath said, "I don't see where any of the proper procedures to seal the documents were ever followed ..." but that he would give Epstein's legal team the ability to "jump through the hoops to seal the documents if they are entitled in fact to be sealed."

The sealing of court documents in Florida is unusual and lawyers typically have to prove a significant reason for it, such as protecting trade secrets or a compelling government interest.

Goldberger said after the hearing there is no rush to unseal the plea deal.

"I think the records clearly need to be sealed and continue to be sealed but I think the ruling by Judge Colbath was a very well- reasoned practical decision," Goldberger said. "He is not getting special treatment."

Brad Edwards, of the law firm of Rothstein Rosenfeldt Adler of Fort Lauderdale, said the plea deal should be a public record. His firm represents the woman, now 20, who was identified only as E.W.

Whether procedure was followed is not the issue, he said.

"Certainly it should be unsealed regardless. I mean this is a very unusual circumstance where a document like this would be sealed," Edwards said. "None of the other criminal defendants in there (Judge Colbath's courtroom) had their plea bargains, plea agreements, their non-prosecution agreements, sealed."

Edwards said his firm represents three women who claim they were procured for sex with Epstein when they were underage. The three are outraged that the document is under wraps, Edwards said.

Deanna Shullman, the attorney representing the Post, said the public and the press have a constitutional right of access to public records in Florida.

"Fortunately, the status quo is openness. So I think the judge has the idea that the initial closure was done without any adherence to those procedures but was inclined to give Mr. Epstein's lawyers additional time to prove that they should be sealed in accordance with these procedures," Shullman said.

"It's a little disappointing in that we would have liked to see the judge unseal the record because that's what should be the status quo in a situation like this," Shullman said.

-- drogers

@pbdailynews.com

LOAD-DATE: September 1, 2010

1 of 11 DOCUMENTS

Copyright 2009 Associated Newspapers Ltd.
All Rights Reserved

The Evening Standard (London)

December 24, 2009 Thursday

LENGTH: 824 words

HEADLINE: CITY SPY

BODY:

EXPECT more media firms to announce plans to charge for content online in early 2010. City Spy hears that business-to-business publisher United Business Media is the latest outfit which is thinking of ramping up its subscription model. Property Week and Building are among the titles which recently started asking users to register their details to keep reading stories, which is seen as a possible precursor to charging.

BUSINESSES TIPPED TO COME A CROPPER

AMID all the contradictory forecasts for recovery or double-dip recession in 2010, what do the insolvency practitioners say? City Spy's mole in the bean-counting world says the last quarter of 2009 was surprisingly quiet as the economy stabilised but they are not optimistic about the new year: "We reckon there's going to be a rush of insolvencies in the second quarter, after the end of the financial year." The next quarterly rent review is due tomorrow, Christmas Day, then again at the end of March. But given the number of "seasonal sales" that started on the High Street at least a week before Christmas, it would be no surprise to see some retailers come a cropper sooner...

EPSTEIN PILOT TAKES TO THE ROAD

FURTHER news reaches City Spy of former Bear Stearns trader, Prince Andrew's shooting companion and convicted sex offender, Jeffrey Epstein.

The ex-Wall Street star served 13 months in jail on criminal charges of soliciting prostitution and procuring a minor for prostitution and he now faces civil claims from young women accusing him of having unlawful sex with them. This week, City Spy recounted how Epstein had transferred the title deeds of his prized 2003 Ferrari 575M Maranello to his private pilot Larry Visoki, prior to the car going on sale for $159,000 (£99,000) (possibly to help Epstein pay his legal bills). It turns out, the same Visoki was deposed last week by Bradley Edwards, an attorney for three of the women suing Epstein. Questioned by Edwards about plane passengers who might have witnessed Epstein in the company of young girls, Visoki admitted Bill Clinton, Prince Andrew, former Israeli prime minister Ehud Barak, former Colombian president Andrés Pastrana Arango, Obama economic adviser Lawrence Summers, billionaire Ron Burkle, and actors Kevin Spacey and Chris Tucker had been on board the plane while young girls were present. Fortuitously for Epstein, however, Ferrari-selling Visoki swore on oath that he never suspected his boss of having sex with them.

Of course not, Larry. Now drive off into the sunset.

More on Prince Andrew, our special representative for international trade and investment. The European Parliament and the Organisation for Security and Co-operation in Europe have strongly condemned Azerbaijan for tightening restrictions on the media and jailing two bloggers who were critical of the government. It transpires the oil-rich country has long blocked BBC broadcasts there, which might explain why oft-criticised Andrew and former Prime Minister Tony Blair spend so much time visiting the sometime Soviet State.

What does the snow have in common with the recession? Every other country can get out of both but Britain can't get out of either. HAPPY news: private jet travel is back, reports the Wall Street Journal. Alas, there is a "but" [#x2039] in-flight food remains in recession. Apparently, those who supply food to executive aircraft are seeing demand soar after a slump, but says one caterer: "No one is eating lobster. A quick turkey box lunch is the order of the day." Of course, that has nothing to do with the industry being desperate to re-brand itself as time-saving and cost-efficient.

WHICH insurance broker saw a compliance officer pass out after the office Christmas lunch and have to be taken to hospital?

WHO MADE OFF WITH THE MONEY? IT's a year since the Bernie Madoff affair blew up and the hedge fund king was found to have been ripping off his clients. If he was in Britain the old fraudster would still be at liberty as lawyers pored over his case and the prosecution had barely cranked into operation. But the US is different [#x2039] his case is done and dusted, and he's languishing in jail. Even so, by US standards, the Madoff conviction was going some. Rumours persist that he pleaded guilty as quickly as he did and said the absolute minimum because he wasn't the main crook of the piece [#x2039] the main business of his hedge fund was washing money for organised crime. As soon as the balloon went up and he was arrested, he was warned by friends with Italian-American origins that his life, and the lives of his family, would be at risk were he not to "take the rap".

OETaking the rap': hedge fund fraud Bernie Madoff

UNFORTUNATE name? City Spy's eye is drawn to a forthcoming lecture at the Institute of Advanced Legal Studies, School of Advanced Study, University of London. It's in partnership with the Market Abuse Association. What? Do they wear a club tie? Do they refer to each other as fellow market abusers?

LOAD-DATE: December 24, 2009

6 of 11 DOCUMENTS

Copyright 2009 ProQuest Information and Learning
All Rights Reserved
ProQuest SuperText
Copyright 2009 Palm Beach Post
Palm Beach Daily News

September 20, 2009 Sunday
Dn1 Edition

SECTION: A SECTION; Pg. A.1

LENGTH: 1126 words

HEADLINE: ATTORNEY FOR EPSTEIN VICTIMS: 'I HAVE NEVER SEEN A STRANGER CASE'

BYLINE: MICHELE DARGAN, MICHELE DARGAN, Daily News Staff Writer

BODY:

Sex offender Jeffrey Epstein could have been charged with multiple counts of five federal offenses involving sex acts with minors and faced a life sentence, but, instead, the government agreed not to prosecute him or his procurers if he spent 18 months in the county jail on two state charges.

Those were the details unsealed Friday in a nine-page federal non- prosecution agreement that lets Epstein and co-conspirators Sarah Kellen, Adriana Ross, Lesley Groff and Nadia Marcinkova off the hook for any of those past crimes.

"He could have gone to prison for life and somehow he's getting immunity in exchange for nothing?" said Fort Lauderdale attorney Brad Edwards, who represents three Epstein victims. "I have never seen a stranger case. To me, it's more spectacular what's not in it. It's the U.S. Attorney's Office saying we'll do everything in our power to see he doesn't get punished."

Edwards has been fighting for a year in federal and state court to unseal the agreement.

"The non-prosecution agreement raises more questions than it answers," said Miami attorney Adam Horowitz, who represents seven victims. "Why did all the co-conspirators receive immunity? Why were the victims not consulted regarding the sentence? Why did he receive such a minimal sentence?

The federal deal has remained sealed in Epstein's state court file since he pleaded guilty in June 2008 to state charges of procuring a minor for prostitution and soliciting prostitution.

U.S. Attorney's Office does not comment

The federal charges he could have faced were: conspiracy to persuade minor females to engage in prostitution, conspiracy to travel to engage in illicit sexual conduct with minor females, persuading minor females to engage in prostitution, traveling to engage in illicit sexual conduct with minor females and causing a person under 18 years to engage in sex for money while knowing they are underage.

The charges carry various statutory penalities ranging from 10 years to life, with a minimum mandatory of at least 10 years.

Alicia Valle, spokeswoman for the U.S. Attorney's Office in Miami, declined comment.

Expert: Feds take few sex-assault cases

North Palm Beach criminal defense attorney Barry Maxwell said he is not surprised that federal charges weren't filed.

"My experience has been that the federal government does not intervene in sex-assault cases, except if we're dealing with a serial rapist or it crosses jurisdictional lines," Maxwell said. "It's either not a big enough case or not atrocious enough for them."

Epstein, 56, served 13 months of his 18-month sentence at the Palm Beach County Stockade and received liberal work-release privileges while in jail. He was able to go to his West Palm Beach office six days a week for up to 16 hours a day.

He is now serving one year of probation at his Palm Beach mansion and is registered as a lifelong sex offender.

Epstein 'fully abided' by deal, says defense

Epstein's attorney Jack Goldberger released the following statement: "This document relates to allegations that were made many years ago. It was by its provisions and agreement of the parties to remain confidential in part to protect the identities of collateral third parties.

"Mr. Epstein has fully abided by all of its terms and conditions. He is looking forward to putting this difficult period of his life behind him. He is continuing his longstanding history of science philanthropy both here in South Florida and nationwide."

Goldberger had blocked the unsealing by filing court papers asking that the documents stay sealed "to prevent a serious imminent threat to the fair, impartial and orderly administration of justice; to protect a compelling government interest; to avoid substantial injury to innocent third parties; and to avoid substantial injury to a party by disclosure of matters protected by a common law and privacy right, not generally inherent in these specific type of proceedings, sought to be closed."

Circuit Judge Jeffrey Colbath ordered the agreement to be unsealed in June, but Epstein's attorneys appealed the ruling to the Fourth District Court of Appeals, which affirmed Colbath's ruling. Colbath had ruled that the federal agreement -- sealed in state court -- was improperly sealed.

'I felt it was my fault'

More than a dozen lawsuits against the billionaire money manager have been filed in federal and state court, all with similar allegations: that a minor girl was taken to Epstein's mansion on El Brillo Way and led upstairs to a spa room by one of Epstein's assistants, where he would ask the girl to perform massages and/or various sex acts, for which he would pay her.

One victim, who is known as Jane Doe #5 in a federal court lawsuit against Epstein, said she didn't find out about the deal until after it was finalized. She was 15 at the time one of her schoolmates told her she could make $200 by giving a massage to a man in Palm Beach.

She says she was "nervous and scared and wanted to leave" once she got to Epstein's spa room.

"I thought, 'I can't call my dad or my mom because I'm stuck in this situation and didn't know what to do," she said. "I really didn't know what this man was capable of. For a long time, I felt like it was my fault and that's exactly what he wanted me to feel."

Epstein has curfew

While he is serving the 12 months of house arrest at his Palm Beach home, Epstein must observe a 10 p.m. to 6 a.m. curfew, have no unsupervised contact with anyone younger than 18 and not view, own or possess pornographic or sexual materials.

The indictment followed an 11-month investigation by Palm Beach police, who said Epstein paid five underage girls for massages and sometimes sex at his El Brillo Way home. Then-State Attorney Barry Krischer declined to prosecute Epstein on multiple charges involving unlawful sex acts with minors. Instead, he brought the case to a grand jury, which charged Epstein on the lesser charge of soliciting prostitution.

Then-Palm Beach Police Chief Michael Reiter wrote Krischer a letter asking him to recuse himself from the case. When that didn't happen, Reiter requested an FBI investigation to determine if any federal laws were broken.

'Out of the ordinary'

West Palm Beach criminal defense attorney Gregg Lerman said several aspects of the Epstein case are unusual.

"I don't understand why it would be a federal case in this circumstance, and why was there anything in writing at all and why did they seal the agreement?" Lerman said. "Why did it go to the grand jury instead of through the state filing lewd assault charges? That's unusual. And it's very unusual that they structure a plea to get county time rather than prison time. That's definitely out of the ordinary. Nobody goes to county jail as a state criminal punishment."

-- mdargan

@pbdailynews.com

GRAPHIC: Caption: Epstein Deal does not allow prosecution of co- conspirators.

LOAD-DATE: September 1, 2010

PalmBeachDailyNews.com
THE SHINY SHEET

Print this page Close

Jeffrey Epstein address book 'Holy Grail' of famous names

By **MICHELE DARGAN**
DAILY NEWS STAFF WRITER

Updated: 7:58 p.m. Friday, March 11, 2011
Posted: 7:57 p.m. Friday, March 11, 2011

When talking about the personal address book of billionaire sex offender Jeffrey Epstein, the term “little black book” takes that phrase to a whole new level.

Manhattan money manager Epstein’s book reads like a laundry list of the world’s richest and most powerful people, including some Palm Beachers.

Referred to as “The Holy Grail” by Epstein’s former house manager — now serving time for trying to sell it to attorneys — the 97-page address book details multiple addresses, phone numbers, e-mails and other contact information for former President Bill Clinton, Britain’s Prince Andrew and Sarah Ferguson, Donald Trump, Sen. John Kerry, various members of the Kennedy clan and former British Prime Minister Tony Blair, among many others.

The British press has been having a field day digging up new details about Epstein’s friendship with Prince Andrew. Virginia Roberts — known only as Jane Doe 102 in court papers — has been dishing her story to London’s Daily Mail. Details include Roberts having been in the company of the prince three times at Epstein’s behest.

Roberts also recounted meeting Clinton on Epstein’s private Caribbean island, according to the Mail.

But Roberts made no suggestion of sexual relations with Prince Andrew or with Clinton, the Mail reported. Similarly, there is no suggestion of anything salacious with any of the Palm Beachers listed among the money manager’s contacts.

Roberts — who spent four years with Epstein — refers to him as “a monster” who paid her lavishly to satisfy his and his friends sexual whims — although Roberts doesn’t identify the friends.

British papers also have reported that Ferguson accepted £15,000 from Epstein. The money was paid to her former assistant, who claimed Ferguson owed him unpaid wages and other bills. Ferguson has since told the Mail and other British papers that she made “a gigantic error of judgment” in accepting the money from Epstein and that she will pay him back.

The entire “Grail” was made public as part of a pending civil court case in which Epstein is suing attorney Brad Edwards, who represented several underage girls who sued Epstein.

Epstein sued Edwards, alleging he was involved in false claims made by Ponzi schemer Scott Rothstein; Edwards countersued Epstein, saying he filed a frivolous lawsuit to get him to back down from representing the young women.

All the lawsuits against Epstein said his modus operandi in the initial visit was the same: The girls were taken to Epstein’s Palm Beach mansion and led upstairs to a spa room by one of Epstein’s assistants, where he would ask the girls to perform sexually charged massages and/or various sex acts, for which he would pay them.

Other high-profile names in Epstein’s book include Special Envoy for Middle East Peace George Mitchell, New York City Mayor Michael Bloomberg, New York Gov. Andrew Cuomo, Barbara Walters, Alec Baldwin, Ralph Fiennes, George Hamilton, Dustin Hoffman, Kevin Spacey, Liz Hurley, Lauren Hutton, Janice Dickinson, Naomi Campbell, Christy Turlington, Henry Kissinger, Joan Rivers, Courtney Love, Mick Jagger, Cornelia Guest, Phil Collins, Itzhak Perlman, Simon LeBon, Charlie Rose, Richard Branson, playwright Candace Bushnell, designers Tom Ford and Vera Wang, soap opera actress Nadia Bjorlin and erotic film star Koo Stark, who once dated Prince Andrew.

Among the high-powered Palm Beachers listed in the money manager’s address book are Catherine and Fred Adler, Samantha and Serena Boardman, Jimmy and Jane Buffett, Pepe Fanjul, Conrad and Barbara Black, Gerry Goldsmith, Marjorie Gubelman, Dana Hammond, David Koch, Henry Kravis, Frayda and George Lindemann Sr., Bob and Todd Meister, Alfred Taubman, Stanley, Bea and Brett Tollman, and Martin Trust.

Gaston Cantens, a spokesman for Florida Crystals Corp., said Fanjul and Epstein “obviously knew each other and had some contact in the past. But there isn’t any ongoing business or social relationship with Mr. Epstein.”

Wexner.”

Les Wexner, CEO of The Limited Brands, was reported to have been Epstein’s biggest client and close friend. Wexner replaced Epstein as his money manager, according to recent reports.

Some of the names in Epstein’s book are sub-listed under geographic locations. The heading “massage” is notated under many of Epstein’s locations. Names and phone numbers, most of them first names only, are listed under the massage entries.

Registered as a level 3 sex offender, Epstein pleaded guilty to soliciting underage girls for sex at his El Brillo Way home in Palm Beach. In addition to serving 13 months of an 18-month jail sentence, Epstein has settled at least two dozen lawsuits with young women for undisclosed amounts.

Find this article at: Print this page Close

http://www.palmbeachdailynews.com/news/jeffrey-epstein-address-book-holy-grail-of-famous-1315130.html

The Telegraph

Jeffrey Epstein 'kept a diary of his under-aged victims'

The Duke of York's billionaire paedophile friend kept a secret journal, described as "The Holy Grail" by lawyers, which listed his alleged under-aged victims and the celebrity guests he entertained at his Florida mansion.

Jeffrey Epstein and Ghislaine Maxwell at Sandringham in 2000 Photo: ALBANPIX

By Jon Swaine (http://www.telegraph.co.uk/journalists/jon-swaine/) , New York

8:00AM GMT 05 Mar 2011

8,591 followers

Jeffrey Epstein used the "black book" to log contact details of the girls that gave massages to him and his friends and those of his powerful and famous associates, such as Bill Clinton and Donald Trump.

A servant at the £4 million manor in Palm Beach, where Prince Andrew enjoyed daily massages during several stays, stole the journal and initially kept it secret from investigators. He is now in prison after attempting to sell it for $50,000 (£31,000).

Epstein, 58, was accused of sex offences by more than 20 under-aged girls. They alleged that after being recruited as masseuses by aides including Ghislaine Maxwell, daughter of the late tycoon Robert, they were seriously assaulted and then paid hundreds of dollars.

The billionaire financier, who attended The Queen's birthday party in 2000, was sentenced to 18 months in prison in 2008, having secured a plea bargain that prevented full criminal trials. He later settled more than a dozen multi-million dollar civil lawsuits from his alleged victims out of court.

The previously undisclosed journal, however, “detailed the full scope and the extent of Epstein's involvement with underage girls”, according to lawyers for several alleged victims.

Prosecutor of the Duke of York’s sex offender friend Jeffrey Epstein speaks of 'assault’ (http://www.telegraph.co.uk/news/uknews/theroyalfamily/8407689/Prosecutor-of-the-Duke-of-Yorks-sex-offender-friend-Jeffrey-Epstein-speaks-of-assault.html)

Duke of York pleads with government over links to dictator (http://www.telegraph.co.uk/news/uknews/theroyalfamily/8363180/Duke-of-York-pleads-for-government-support-over-dinner-with-Tunisian-dictators-relation.html)

It contained the names of girls that Epstein allegedly abused in “Michigan, California, West Palm Beach, New York, New Mexico, and Paris”, according to court papers.

It also listed extensive contact details for Epstein’s house guests, who had “no connection whatsoever” with alleged offences, including Mr Clinton, the former US President, and Mr Trump, the famous businessman.

It could not be confirmed last night whether the book contained contact details for the Duke. One lawyer for Epstein’s alleged victims said: “I would bet he is, because he is that good a friend”.

There is no suggestion that the Duke did anything wrong. However he has this week faced a string of questions about his judgment due to his association with a convicted child sex offender.

Epstein kept 21 different phone numbers for Mr Clinton, including some for his assistant and also “Clinton’s personal numbers”, the court papers state.

Alfredo Rodriguez, a servant who “saw numerous underage girls coming into Epstein’s mansion for purported ‘massages’,” took the journal and did not mention it to investigators.

“Because of the importance of the information in the journal to the civil cases, Mr Rodriguez called it ‘The Holy Grail’,” the court documents state.

He attempted to sell the book to lawyers for $50,000 (£31,000) in August 2009 - by which time Epstein had already been sentenced and released from jail.

The lawyers contacted the FBI, who sent an undercover agent to buy the book from Rodriguez. He was then arrested and charged with obstruction of justice.

Rodriguez, who said in sworn testimony that he was terrified of Miss Maxwell, also said he took the book as an "insurance policy" to prevent Epstein making him "disappear".

He was sentenced to 18 months in prison – the same punishment Epstein received for his sex offences – in June last year. His wife, Patricia Dunn, told The Daily Telegraph that he was "sorry" for his error.

Dave Lee Brannon, an Assistant Public Defender who represented Rodriguez, said: "If this book had been produced when requested, Mr Epstein's sentence may have been significantly different."

Details of the journal emerged in a lawsuit brought in Florida by Epstein against Brad Edwards, a lawyer representing several alleged victims.

Epstein alleges Edwards was linked to a "Ponzi scheme" run by a colleague, which lured investors by falsely claiming Epstein had agreed to settle sex-offence lawsuits for hundreds of millions of dollars.

Edwards rejects the allegation, which has already been dismissed by the Florida Bar.

© Copyright of Telegraph Media Group Limited 2011

PalmBeachDailyNews.com
THE SHINY SHEET

Judge denies gag order in Epstein, Edwards lawsuit; dismisses complaint

By **MICHELE DARGAN**
DAILY NEWS STAFF WRITER

Updated: 7:25 p.m. Wednesday, July 13, 2011
Posted: 7:11 p.m. Wednesday, July 13, 2011

A circuit judge Wednesday squashed an attempt by attorneys for Jeffrey Epstein to prevent parties in a civil lawsuit involving the billionaire sex offender from talking to the media.

Circuit Judge David Crow denied the motion for a gag order.

In addition, Crow dismissed the complaint by Epstein against attorney Brad Edwards, who has represented 10 underage girls in sex abuse claims brought against Epstein. Epstein alleged Edwards abused the court system by threatening to depose Epstein's powerful friends, which included Donald Trump and President Bill Clinton.

Other claims included that Edwards tried to obtain records from an alleged sex therapist who had never treated Epstein and that Edwards used investigative tools that included trespassing on Epstein's property.

Crow gave Epstein's attorneys 30 days to refile the lawsuit, which will be the second amended complaint and third evolution of the lawsuit.

Epstein also named Edwards' former boss, convicted Ponzi schemer Scott Rothstein, in the lawsuit, alleging Rothstein made "various representations to potential investors regarding the Epstein actions."

In dismissing the lawsuit, Crow said he found "serious problems" with the complaint. Crow called the lawsuit "vague.

"You have to know, at this point in time, what he did or didn't do that was an abuse of process," Crow said.

Epstein, 58, has confidentially settled more than two dozen lawsuits with young women who allege they were sexually abused by him when they were minors.

Edwards filed a counterclaim, alleging Epstein filed the lawsuit to get Edwards to back down from representing the victims.

"Mr. Epstein had to pay more to settle these cases than he would have if Mr. Edwards wasn't out there putting all this pressure on him," said attorney Jack Scarola, who represents Edwards. "That's Mr. Edwards' job ... to put as much legitimate pressure on the defendant as he possibly could and he obviously did an extremely effective job."

Epstein pleaded guilty to two felony charges: soliciting prostitution and soliciting a minor for prostitution. He served 13 months in the county jail and has to register as a lifelong sex offender.

Representing Epstein, attorney Joseph Ackerman argued unsuccessfully for a gag order. Ackerman said Scarola has repeatedly made statements to several news organizations about the case.

"Mr. Scarola has constantly referred to Mr. Epstein as a pedophile and there's been no proof of that anywhere," Ackerman said. "Muzzling lawyers who may wish to make public statements has been long recognized as within the court's inherent power ... We don't believe it's appropriate to wage a media campaign and taint the jury pool."

Scarola said it would be unconstitutional to impose a gag order.

"There is a complete and total absence of proof that we have engaged in any conduct whatsoever that could be prohibited," Scarola said.

Scarola said he and Edwards have been asked to appear on national television as well as received interview requests from the foreign press. Scarola said he has been selective in his interviews.

There has been a frenzy in the British press ever since Virginia Roberts, an Epstein victim, spoke to the Sunday Daily Mail earlier this year about being introduced to Prince Andrew and spending time with him at Epstein's behest. Roberts alleged she served as a sex slave to Epstein when she was a minor.

“I’m pleased to hear he’s embarrassed by his conduct. Maybe it will serve as some deterrent in the future.”

Find this article at:

Print this page Close

http://www.palmbeachdailynews.com/news/judge-denies-gag-order-in-epstein-edwards-lawsuit-1606397.html

Navigating our presidential campaign was a piece of cake compared to understanding the nuances of the 2011 Oscar race for the most revered artistic honor in the world.

This is how nine films fell into the big picture.

Three premiered in Cannes mid-May, a distant nine months ago, creating an Oscar campaign as long as any human pregnancy.

At the Palais, the first inkling of Oscar buzz was born as reclusive Woody Allen premiered "Midnight in Paris".

PBS later aired a documentary of Woody discussing his forty-four films showing the astonishing depth of his talent that made you want to immediately hand him the Oscar for Best Picture. Academy rules and Woody forbade marketing this gem.

Woody is not a member of the Academy because he doesn't feel that films should be in competition. He told me, "A statue does not change your life. You still get a cold. You can't get a date. You still have everyday things to worry about". The Academy learned to love him from a distance and gave him best original screenplay as a consolation.

Terrence Malick's long awaited esoteric "The Tree of Life" unveiled at Cannes and won the coveted " Palme D'Or " positioning it for a nomination.

"The Artist", created by the French, shot in Hollywood and about Hollywood was the festival surprise. This charming and oddly original black and white silent entry was introduced by the ringmaster himself, Harvey Weinstein. No one could pronounce or spell director Michel Hazanavicius's name. Jean Dujardin could not speak a word of English and neither could his 10 year old co-star, the Jack Russell Uggie who had been rescued from the pound after two adopters found him too wild. Tragically Uggie developed an undisclosed neurological disorder during production, forcing him to retire at the height of his popularity.

No slam dunk Oscar winner emerged in Cannes. Any future film could easily win.

DreamWorks' "The Help" premiered in LA in August and distributer Disney began propelling the politically correct and socially significant film to box office heaven of $200 million. Viola Davis and Octavia Spencer were forecast to win Oscars.

In September, the Toronto and New York Film Festivals and Fox Searchlight presented Alexander s's Hawaiian family saga, "The Descendants," which broke out of the pack with whispers of winning. Beloved George Clooney, playing a father for the first time was hailed a shoe-in for best actor. Directing "Ides of March" was additional momentum.

Also at New York's festival Marty Scorsese and Paramount sneaked an unfinished cut of "Hugo" in Alice Tully Hall, built for concerts but converted into a 3-D theater. Marty was christened the visionary genius of an innovative costly 3-D masterpiece.

Director Bennett Miller's highly anticipated "Moneyball" for Sony hit a grand slam at its west coast premiere in Oakland aligning the film, it's heart throb star Brad Pitt, Jonah Hill and seasoned writers Steve Zaillian and Aaron Sorkin in play.

Spielberg's epic "War Horse" for DreamWorks came thundering down the pike with a huge premiere back at Alice Tully Hall with posters of Lincoln Center's Tony winning theatrical "War Horse" with their indelible puppets in the background. Steven paid homage to legends John Ford and David Lean and the country fell in love with a horse named Joey and his fourteen stand-ins.

Studios worked their stars to the bone. Ironically, Harvey's independent French talent who lived in Paris were not as available as their competitors, therefore Uggie became a super star igniting a pet war.

Christopher Plummer, who had best supporting actor in the bag promoted his Jack Russell, Cosmo. Diminutive Scorsese was seen on TV on a small couch with his large Doberman, Blackie drooling on his suit. Spielberg never got a chance to trot out his lead horse Joey, previously seen in "Seabiscuit" because his ravishing reddish coat was now darkened for another role.

By December, as film critics bestowed their own awards upon many films, Stephen Daldry struggled to finish "Extremely Loud and Incredibly Close" with a new score. There was buzz Daldry could be editing the winner. Producer Scott Rudin juggled his astounding three films in one year from Daldry, Miller and David Fincher directing “The Girl with the Dragon Tattoo.”

Daldry had received three consecutive directing nominations. In January, for his fourth film, he received a best picture nomination, for a boy's emotional journey dealing with 9/11, and the nine films were officially off and running. Forty-five film and media groups handed out awards leading up to Oscar night.

Wednesday, February 22

My airplane seatmate to LAX was Sony Classics Michael Barker. The night before Woody Allen had shown Michael "Nero Fiddled", his **latest** film shot in Rome rumored to be his best.

When Woody won the Oscar Sunday night, for a record breaking 23rd overall nomination, he had just finished pasta a Sette Mezzo on Lexington Avenue with art dealer Lorinda Ash and Soon-Yi. He went home and watched the N.B.A. All-Star game. Soon-Yi watched the awards show on a delay in another room. By the time Woody won, he fell asleep and Soon-Yi didn’t want to wake him. The next morning he went to the breakfast table alone and read in The New York Times he won. He had to think it was a good omen and he would not catch a cold that day.

Before Michael and I flattened our recliner chairs for the big sleep, I told him I felt confident his Iranian film, "A Separation" was winning best foreign picture. He told me "The Artist" would take best picture and director. Actor was a tight race between Brad Pitt, Jean Dujardin, the "Clooney of France" and the real George Clooney. George was essentially running against a version of himself, which only slightly amused him.

The biggest dilemma was Viola vs. Meryl. Michael picked Meryl as New Yorkers did. "The Help" had taken on a life of its own lead by vivacious Viola in LA. "The Iron Lady", a much criticized film showcased Meryl's tour de force performance. Few knew at the last minute, on President's weekend Harvey's shout out, "She hasn't won in 29 years!" resonated.

An androgynous driver named Monica greeted me at the airport in a black tuxedo that would make Albert Nobbs weep for joy, prompting me to devilishly think of her as "Nobbs" all weekend. She barely recognized me sporting a new Sally Hershberger hairdo, "the yenta with the dragon tattoo."

Checking into the Beverly Hills Hotel I bumped into best actress nominee, Golden Globe and Spirit Award winner Michelle Williams with her daughter Matilda Ledger headed to the swimming pool. Innocently standing there with no makeup she was remarkably the antithesis of Marilyn Monroe. I told her she so deserved the Oscar for her mesmerizing transformation which did not cheer her up knowing the gold was going to Viola or Meryl.

This year there seemed to be more parties than ever. Vanity Fair publisher Edward Menicheschi staged a staggering six nights of “CAMPAIGN HOLLYWOOD.” Ermenegildo Zenga and Colin and Livia Firth hosted an intimate dinner at the Chateau Marmont to benefit Oxfam America, Colin’s pet charity. Editor Graydon Carter and Edward greeted Cameron Diaz, Kristin Davis, Gary and Alexandra Oldman, Mia Wasikowska. In addition, Livia spoke about her 'Green Carpet Challenge' which uses eco-friendly fabrics for “wear it once” gowns at awards shows. Get it? Go green on red.

It was a busy night at the Chateau as French billionaire investor Nicolas Berggruen hosted his party for young, leggy, breathtakingly beautiful models. Lindsay Lohan was on the loose after a judge having announced she was making great progress. She showed up sober wearing a new face bloated with injectable Restalyne that could only be preparation to portray Elizabeth Taylor in "Liz and Dick", her film for Lifetime. Peter Brant and Stephanie Seymour and Peter Morton and Linda Evangelista joined Josh Harnett and Emile Hirsh under the terrace heaters. The boys were looking for love in all the right places.

Thursday, February 23

Thursday night boasted fifteen events causing party panic. Here is a brief rundown of **seven**.

At The Hollywood Reporters Nominees Night, editor Janice Min and publisher Lynne Segall greeted the power brokers. With ballots in, competing studios cordially **mingled in the mayor’s backyard.** Owen Wilson slipped in the back door and hung with Michael Sheen and producer Letty Aronson. Producers Kathy Kennedy and Frank Marshall chatted with DreamWorks **partner Stacey Snyder, producer** Graham King **and** Emily Mortimer. Fox's Tom Rothman and Jim Gianopulos compared notes with Focus's James Shamus. Young directors Drake Doremus ("Like Crazy"), Sean Durkin ("Martha Marcy May Marlene") and Oscar nominee and Spirit Award winner J.C. Chandor ("Margin Call") **drank at the bar**. Breakout directors Nick Jarecki ("Arbitrage"), Zal Batmanglij ("Sound of My Voice and "The East") and Jay Duplass ("Jeff Who Lives at Home") dreamed about their future nominations. Aaron Sorkin, Piers Morgan and Lawrence O'Donnell handicapped Romney vs. Obama as Brooklyn Decker sashayed by.

Urs Fisher's exhibition Beds & Problem Paintings featured two bed sculptures at Larry Gagosian's Gallery followed by his private dinner at Mr. Chow's. Art lovers Vera Wang, Russell Simmons, Steve Martin, Jean Pigozzi and John Waters attended.

The US-Ireland Alliance honored nominees "Hugo" screenwriter John Logan, "Bridesmaids" star Melissa McCarthy and Michelle Williams at Bad Robot**.** Logan also wrote "Rango” "Coriolanus", 007’s "Skyfall" and "Jersey Boys" for the big screen.

Alfre Woodard hosted a down and dirty girls night out in a rented house above Sunset for Viola Davis and Octavia Spencer.

Universal honcho Ron Meyer hosted a civilized private buffet at his Malibu home for Graydon Carter with Tom Cruise, Leonardo DiCaprio, Barbra Streisand, Tom Hanks and Michael Douglas.

Stunning socialite Betsey Bloomingdale gave a seated dinner at her Holmby Hills home for best friends Nancy Reagan, Wendy Stark, Bob Colacello, Joan Collins and fashion icon Lynn Wyatt.

Tobias Meyer, **auctioneer** for Sotheby's and art dealer Mark Fletcher hosted an open house at their Mulholland Drive home for English **avant garde** photographer Terry Richardson. This is the only pre-Oscar party where a guest dropped his pants and mooned the red carpet and Terry signed a fans breast. Art collectors Bill and Maria Bell, Todd Eberle, rock singer Jack Donahue **and** Francesco Clemente ____schmoozed.

Friday, February 24

At the BHH I ran into David Heyman, English producer of the "Harry Potter" franchise who was honored at the Publicists Awards lunch at The Beverly Hilton. "Motion Picture Showman of the Year" was the consolation prize for being snubbed by the Academy for visualizing a publishing miracle for children around the world.

"Nobbs" whisked me off to the British Film Reception hosted by Jeremy Hunt, UK Secretary for **Culture and Olympics,** and the British Consul-General Dame Barbara Hay, in her Hancock Park residence. Upon introduction, I blurted out that my friend Lord Astor, was interested in having LA people get to know his son-in-law, Prime Minister David Cameron. As an appointed diplomat she was horrified by my indiscretion and turned to greet the next American idiot. I was just making conversation.

Daldry told Sony's Sir Howard Stringer and astute film CEO Michael Lynton, Kenneth Branagh, Janet McTeer, **and** Gary Oldman that he, as executive producer of the Olympics, was headed back to London to oversee special events, including the opening ceremony, directed by Danny Boyle.

Victoria Beckham made a dramatic sullen last-minute appearance looking perfectly skinny in her own designed dress.

At the Women in Film party at Cecconi's, Gwyneth Paltrow, Shailene Woodley, Selena Gomez and Vanessa Hudgens networked with Jessica, Octavia and Viola now of social stamina fame.

Blythe Danner kissed me at the door as a military type looked on. I kept saying to him, "Where have we met?" Nowhere. He was astronaut Mark E. Kelly who came with Blythe and is married to former congresswoman Gabrielle Giffords, Paltrow's second cousin. Only I could mistake an astronaut for a movie marketing guy.

Vanity Fair feted Scorsese and The Film Foundation which has saved 555 films in 22 years. Cocktails were at the restored Bel-Air Hotel. Honorary Jewess Lorraine Bracco ran past me yelling she was late for Ronald Perelman's Shabbat dinner. Three, three-time Oscar winners: composer Howard Shore, costume designer Sandy Powell and editor Thelma **Schoonmaker** were honored. Sir Ben Kingsley, Danny Huston, Patty Clarkson, Irwin Winkler and Giorgio Armani's niece Roberta Armani with Wanda McDaniels debated best director: Marty or Michel?

“Nobbs” delivered me to WME's Party at kahuna Ari Emanuel's Brentwood estate, where NFL quarterback and new client Tim Tebow was the toast of the party, especially to Taylor Swift who made $35.7 million this year. Michael Douglas gave me a kiss…**doesn’t get any better**. Lovebirds Robert Pattinson and Kristen Stewart made a rare appearance, glued to each other's hips. They mingled with co-star Taylor Lautner, Miley Cyru**s** and her "Hunger Games" beau Liam Hemsworth. Meanwhile Charlize Theron, Jack Black, Rooney Mara, Ben Stiller, Barry Sonnenfeld and Larry David talked business with moguls Les Moonves and Viacom's Philippe Dauman.

Next was UTA Chairman Jim Berkus' soiree that police almost shut down because the DJ got carried away impressing Harrison Ford, Channing Tatum, Jerry Bruckheimer, Tom Freston, Disney's Rich Ross, SNL's Lorne **Michaels** and Oscar show producer Brian Grazer.

Sunset Tower Hotel owner Jeff Klein and producer John Goldwyn hosted a secretive dinner for Anna and Graydon Carter at their Hollywood Hills home with Tom Ford, Mitch Glazer, Fran Lebowitz, Vito Schnabel, Denise Hale, Lisa and Eric Eisner and VF's Punch Hutton, who is Tim Hutton's sister. _____

Last stop was CAA Byran Lourd's "Friday Night Party". “Nobbs” was instructed to drop me off at a neighborhood school where a luxury van transported guests to the stone and glass Bel-Air estate situated on a narrow street. I knew that guests Colin Firth, Penelope Cruz, Sofia Vergara, Salma Hayek, Sandra Bullock, and Glee's Matthew Morrison did not arrive by bus. Once inside the playing field leveled out and pound for pound there was more famous flesh per square inch than **at** the Oscars themselves.

I huddled with Meryl on the couch and we talked about her race. She thought Viola. I thought Meryl. She didn't know about **Harvey’s** last minute "29 year" shout out.

I hugged Bette Midler, flirted with Jim Sturgess and Bennett Miller. Universal's Donna Langley, who is overseeing Tom Hooper's production "Les Miserable" mentioned Hugh Jackman's impeccable manners should insure best behavior from Russell Crowe.

I introduced HBO’s "Game Change" director Jay Roach who is **an** authority on Hitler, to George Clooney who is writing a thriller about the Nazis stealing art. Clooney whispered, “The Frenchman is winning.”

I thanked Bryan Lourd, Kevin Huvane and Richard Lovett, got on the bus and prayed that I get invited back.

Saturday, February 25

I dragged my tired ass to the Academy, as foreign films aficionado Mark Johnson was conducting a symposium on the nominated films, which included Sony Classic’s “In Darkness” from Poland, "Footnote" from Israel and "A Separation" from Iran. Israeli and Iranian governments from the other side of the world monitored their directors as the Sony boys kept the peace.

Michael and Tom Bernard invited me to the Independent Spirit Awards at the Santa Monica Pier. Michael hosted his "Take Shelter" nominees Jessica Chastain, Michael Shannon and director Jeff Nichols. Tom held **court** at the next table with the entire Iranian cast of "A Separation", which won. Tom almost had a heart attack when I threw my arms around Iranian director Asghar Farhadi, which in Iran is unacceptable behavior especially by a Jewish American Princess. Tom wished I had gone to the crowded champagne brunch in honor of Prince Albert and Princess Charlene at the Bel-Air Hotel where Montblanc launched Grace Kelly watches. Simultaneously, TV producer Gary Pudney hosted another secretive lunch which Albert and Charlene actually attended with Graydon Carter, Carolina Herrera, Wallace Annenberg, Bobby Shriver, Bobby Marx, Kathy and Ricky Hilton and Lynn Wyatt. Wolfgang Puck joined them for dessert.

"The French," as the "The Artist" gang was nicknamed, had won six Cesars, France's version of the Oscars in Paris the night before. They flew all night and Harvey's chauffeur arranged a police escort from LAX just in time for them to win four Spirit Awards, cementing the Oscar win.

Back at the BHH, Spielberg was the first to arrive at the tenth annual "Night Before" fundraiser in support of the Motion Picture and Television Fund. Jeffrey Katzenberg had already secured $200 of a $350 million fundraising goal that included money from him, Tom Cruise, Steve Bing, Casey Wasserman, Clooney and Spielberg. Every nominee showed up.

Chanel and Charles Finch cooked up their chic soiree at Madeo, where a mariachi band enthusiastically announced everyone's arrival. Bedecked exclusively in Chanel were Diane Kruger, Rose Byrne, Ginnifer Goodwin, **and** Rachel Bilson . Also air kissing were Rachel Zoe, Rosanna Arquette, Alice Evans and Ioan Gruffudd, Zachary Quinto, and Julia Ormond.

Elizabeth Olsen who has six upcoming films, gushed to Dustin Hoffman that "Wag the Dog" was her favorite movie. I sat next to director, Michael Apted who was editing “Of Men and Mavericks,” a surfer film he co-directed with Curtis Hanson.

“My Week With Marilyn’s” English director Simon Curtis insisted we join Kenneth Branagh at The Weinstein Company’s bash at the Soho House in time to hear Tony Bennett sing “Autumn Leaves” to Harvey’s surprised 86 year-old mother Miriam, Madonna and Meryl. Jean Dujardin, Berenice Bejo, Michel Hazanavicius and producer Thomas Langmann staggered around completely jet lagged **and** too tired to speak English . Uggie on the other hand was the absolute star of the evening as his trainer placed him in everyone’s arms for photo two shots. “W.E’s” Andrea Riseborough and Abbie Cornish sat next to baseball cap-clad Leonardo DiCaprio as his ex Bar Refaeli kept her distance across the room. Zoe Saldana walked in holding hands with Bradley Cooper. Scarlett Johansson introduced me to her boyfriend Nate Naylor. Katy Perry, Felicity Jones and Malin Akerman circled a refreshed Gerard Butler. The two daughters of New York slain hero cop Peter Figoski, Corrine, 14 and Caroline, 16 who had also been Harvey’s guests at the Super Bowl stood in the middle of this circus and just fainted.

Sunday, February 26

I met interior **designer** Nicky Haslam in the Polo Lounge and found Nancy Reagan brunching with Bob Colacello and Carolina Herrera. Nancy knows me as “The DVD Lady**.**” I promised to send her “The Iron Lady”. I didn’t have the heart to tell her that her husband, who was Margaret Thatcher’s best friend, was barely mentioned in the film.

In a Marchesa gown, Dennis Basso fur and Iradj Moini necklace, I collected Simon Curtis and headed to the Hollywood & Highland Center. Simon, who has never been to the Oscars before, miraculously scored a front row seat between Michelle **Williams** and Clooney. We pulled up to screams of hysteria at the mother of all red carpets. We ended up in front of thousands of cameras and I instructed Simon to take baby steps for an hour.

Meryl's publicist Leslee Dart whispered to me. "She is dressed like an Oscar. What do we do if she loses?" Sacha Baron Cohen hilariously guilted the Academy into letting him wear his costume from "The Dictator" and after soiling Ryan Seacrest he went directly to dinner at Vanity Fair. Gwyeneth Paltrow won the style award in Tom Ford's white column and Angelina Jolie so successfully invented a new one legged pose in a thigh high slit gown on the red carpet she repeated it for 39.3 million people onstage. Tickets were so tight that I gave my plus one to Penelope Ann Miller because she promoted "The Artist" every day for four months. Seated next to us were co-stars James Cromwell and Missi Pyle.

Billy Crystal made us laugh, Cirque de Soleil made us gasp and most of the wins were expected.
Colin Firth crowned Meryl her third win on a record 17 nominations and Harvey beamed. ______

The Academy was so confident "The Artist" would **win;** they invited Uggie, who waited in the wings and ran out as Tom Cruise announced the film.

Once the show was off the air, I followed my seatmates to the stage loaded with their programs, wraps and handbags and led the French to the press rooms.

The Governor's Ball was the next stop where the winners got their Oscars engraved. Everyone paid respects to the Academy's Tom Sherak and Dawn Hudson. Stars headed to the Sunset Tower to Graydon's glittering Vanity Fair Oscar bash and their militarized security with micro chipped cards. If your name was not on the list and you were carrying an Oscar, you could walk in. Billy and Janice Crystal were mobbed with **well-wishers**. Elton John made $5 million dollars at his 20th AmFar event which also auctioned off two tickets to Vanity Fair's party for $230 thousand dollars. Tom Cruise and Katie Holmes were thrilled to talk to George Lucas. Gwyneth snuggled with Coldplay's Chris Martin and Jennifer Lopez brought boy-toy Casper Smart. Every major actor previously mentioned was standing in the room. In a sea of celebrity were Tina Fey, David Beckham, Glenn Close, Terry George, Demian Bichir, Claire Danes, Meryl's daughters Mamie Gummer with **husband** Ben Walker**,** Grace Gummer, Salma Hayek, Brit Marling, Natalie Portman, Sofia Coppola, Ryan Kavanaugh, Ingrid Sischy, Sandy Brant and Wendi Murdoch with Bingbing Li.

George Clooney threw his own exclusive party at Craig's in West Hollywood for close friends Bryan Lourd, Grant Heslov, Stan Rosenfield, Brad and Angelina, Emily Blunt and John Krasinski, Cindy Crawford, Jimmy Kimmel, Ryan Seacrest and best adapted screenplay winner Alexander Payne. As much as George supported "The Descendants" he and Brad seemed to have cancelled each other out this year with two great performances.

"The French" and their closest 200 threw their own wild celebration at the Chateau, poured champagne down their throats and threw each other in the pool at 3:00am. Nobody spoke a word of English.

By 4:00am Harvey rounded his talent up for a live broadcast on "The Today Show" from The Four Seasons lobby. In disheveled black tie five Oscar winners and Berenice Bejo, who were total unknowns a year ago made Oscar history with the first silent film to win since 1927. This had to be the most exciting night of their lives. Tomorrow they all go back to **reality** but the glory and the memories will live on forever. As an Oscar winner, **who cares about a cold?**

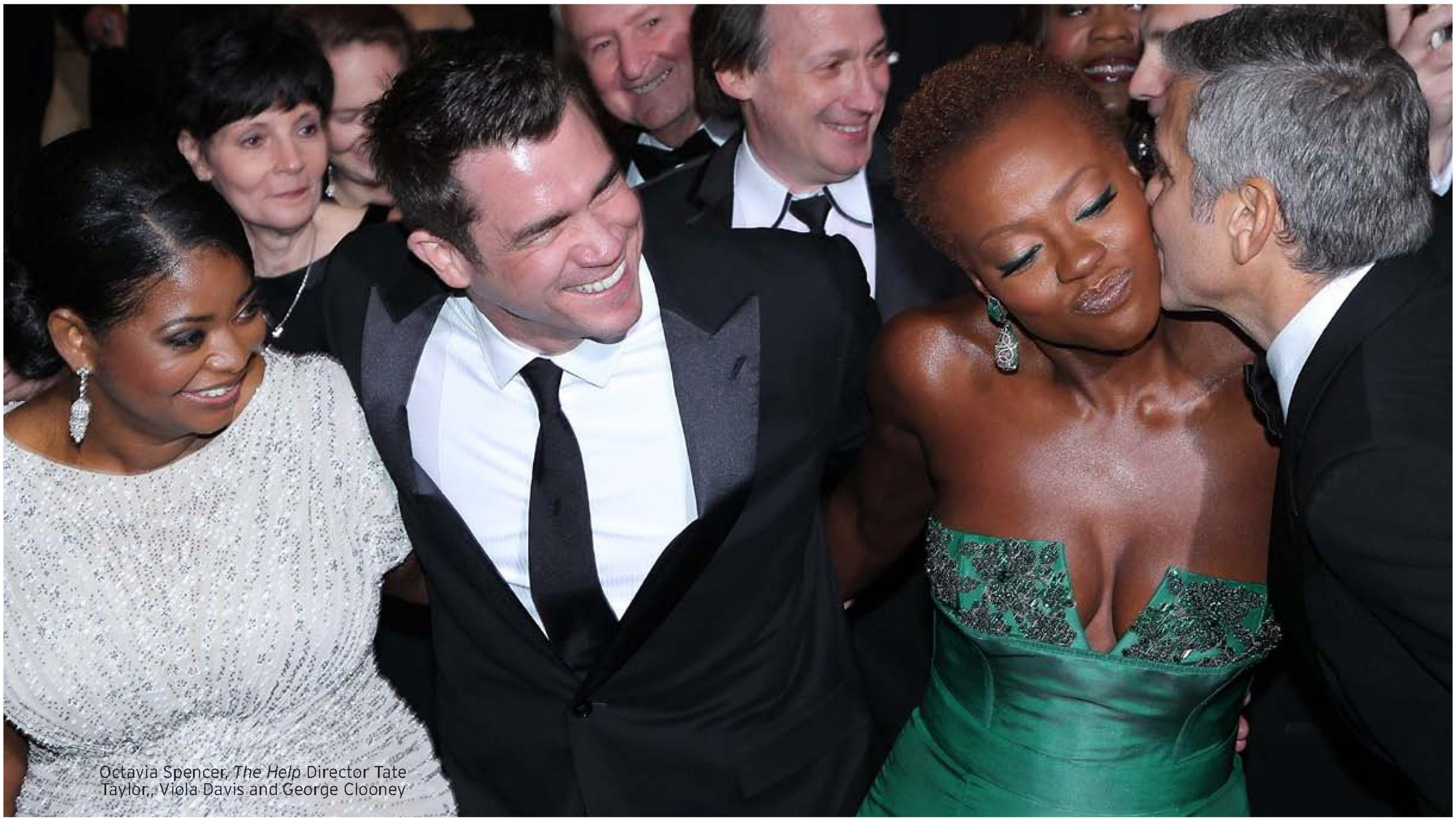
Uggie was invited to the White House correspondent’s dinner as a guest of the Washington Times in April. He hopes to meet President Obama.

VIVE L'OSCARS

Winner Woody Allen may have missed Hollywood's biggest lovefest, but intrepid über movie publicist Peggy Siegal was there for every single party and every single step of the red carpet way. This year, her exclusive Oscar diary chronicles close encounters with Michelle Williams, Meryl Streep, Harvey Weinstein, Nancy Reagan, Elizabeth Olsen, George Clooney and his French doppelganger Jean Dujardin and, of course, Uggie.

photographs by Patrick McMullan *and* Billy Farrell Agency

Katie Holmes and Tom Cruise



Octavia Spencer, *The Help* Director Tate Taylor,, Viola Davis and George Clooney

Navigating our presidential campaign was a piece of cake compared to understanding the nuances of the 2011 Oscar race for the most revered artistic honor in the world.

This is how nine films fell into the big picture.

Three premiered in Cannes mid-May, a distant nine months ago, creating an Oscar campaign as long as any human pregnancy. At the Palais, the first inkling of Oscar buzz was born as the reclusive **Woody Allen** premiered *Midnight in Paris*. PBS later aired a documentary of Woody discussing his forty-four films showing the astonishing depth of his talent that made you want to immediately hand him the Oscar for Best Picture. Academy rules and Woody forbade marketing this gem.

Woody is not a member of the Academy becase he doesn't feel that films should be in competition. He told me, "A statue does not change your life. You still get a cold. You can't get a date. You still have everyday things to worry about." The Academy learned to love him from a distance and gave him Best Original Screenplay as a consolation.

Terrence Malick's long-awaited, esoteric *The Tree of Life* was unveiled at Cannes and won the coveted Palme D'Or, positioning it for a nomination.

The Artist, created by the French, shot in Hollywood and about Hollywood was the festival surprise. This charming and oddly original black-and-white silent entry was introduced by the ringmaster himself, **Harvey Weinstein**. No one could pronounce or spell director **Michel Hazanavicius**'s name. **Jean Dujardin** could not speak a word of English and neither could his 10-year-old co-star, the Jack Russell **Uggie** who had been rescued from the pound after two adopters found him too wild. Tragically Uggie developed an undisclosed neurological disorder during production, forcing him to retire at the height of his popularity.

No slam dunk Oscar winner emerged in Cannes. Any future film could easily win.

DreamWorks' *The Help* premiered in LA in August and distributor Disney began propelling the politically correct and socially significant film to box office heaven of $200 million. **Viola Davis** and **Octavia Spencer** were forecast to win Oscars.

In September, the Toronto and New York Film Festivals and Fox Searchlight presented **Alexander Payne**'s Hawaiian family saga, *The Descendants*, which broke out of the pack with whispers of winning. Beloved **George Clooney**, playing a father for the first time was hailed as a shoo-in for best actor. Directing *Ides of March* added momentum.

Also at New York's festival **Marty Scorsese** and Paramount sneaked an unfinished cut of *Hugo* in Alice Tully Hall, built for concerts but converted into a 3-D theater. Marty was christened the visionary genius of an innovative, costly 3-D masterpiece.

Director **Bennett Miller**'s highly anticipated *Moneyball* for Sony hit a grand slam at its west coast premiere in

Oakland putting the film, its heart throb star **Brad Pitt**, **Jonah Hill** and seasoned writers **Steve Zaillian** and **Aaron Sorkin** in play.

Spielberg's epic *War Horse* for DreamWorks came thundering down the pike with a huge premiere back at Alice Tully Hall, with posters of Lincoln Center's Tony winning theatrical version and their indelible puppets in the background. Steven paid homage to legends **John Ford** and **David Lean** and the country fell in love with a horse named **Joey** and his 14 stand-ins.

Studios worked their stars to the bone. Ironically, Harvey Weinstein's independent French talent who lived in Paris were not as available as their competitors, therefore Uggie became a superstar igniting a pet war.

Christopher Plummer, who had Best Supporting Actor in the bag promoted his Jack Russell, **Cosmo**. Diminutive Scorsese was seen on TV on a small couch with his large Doberman, **Blackie**, drooling on his suit. Spielberg never got a chance to trot out his lead horse Joey, previously seen in *Seabiscuit* because his ravishing reddish coat was now darkened for another role.

By December, as film critics bestowed their own awards upon many films, **Stephen Daldry** struggled to finish *Extremely Loud and Incredibly Close* with a new score. There was buzz Daldry could be editing the winner. Producer **Scott Rudin** juggled his astounding three films in one year from Daldry, Miller and **David Fincher** directing *Girl with the Dragon Tattoo.*

Daldry has received three consecutive directing nominations. In January, for his fourth film, he received a Best Picture nomination, for a boy's emotional journey dealing with 9/11, and the nine films were officially off and running. Forty-five film and media groups handed out awards leading up to Oscar night.

WEDNESDAY, FEBRUARY 22

My airplane seatmate to LAX was Sony Classics **Michael Barker**. The night before Woody Allen had shown Michael *To Rome With Love,* his new film shot in Rome and rumored to be his best. When Woody won the Oscar Sunday night, for a record breaking 23rd overall nomination, he had just finished pasta at *Sette* Mezzo on Lexington Avenue with art dealer **Lorinda Ash** and **Soon-Yi**. He went home and watched the N.B.A. All-Star game. Soon-Yi watched the awards show on a TiVo delay in another room.

Martin Scorcese

Meryl Streep

Lily Collins

Berenice Bejo
and Michel
Hazanaviciu

Michelle
Williams

Princess Charlene and
Prince Albert II
of Monaco

Brad Pitt and
Angelina Jolie

David and
Victoria
Beckham

Jane Fonda and Richard Perry

Scarlett Johansson

Salma Hayek

Billy Crystal

Tom Hanks and Rita Wilson

Jessica Chastain

Emma Stone

By the time Woody won, he had fallen asleep and Soon-Yi didn't want to wake him. The next morning he went to the breakfast table alone and read in *The New York Times* that he had won. He had to think it was a good omen and he would not catch a cold that day.

Before Michael and I flattened our recliner chairs for the big sleep, I told him I felt confident his Iranian film *A Separation* would win Best Foreign Film. He told me *The Artist* would take Best Picture and Director. Actor was a tight race between **Brad Pitt, Jean Dujardin**, the "Clooney of France" and the real **George Clooney**. George was essentially running against a version of himself, which only slightly amused him.

The biggest dilemma was Viola vs. Meryl. Michael picked Meryl as New Yorkers did.

Meryl's publicist Leslee Dart whispered to me. "She is dressed like an Oscar. What do we do if she loses?"

The Help had taken on a life of its own lead by vivacious Viola in LA. *The Iron Lady*, a much criticized film, showcased Meryl's tour-de-force performance. Few knew at the last minute, on President's weekend, Harvey's shout out, "She hasn't won in 29 years!" resonated.

An androgynous driver named Monica greeted me at the airport in a black tuxedo that would make Albert Nobbs weep for joy, prompting me to devilishly think of her as "Nobbs" all weekend. She barely recognized me sporting a new **Sally Hershberger** hairdo, "the yenta with the dragon tattoo."

Checking into the Beverly Hills Hotel I bumped into Best Actress nominee, Golden Globe and Spirit Award winner **Michelle Williams** with her daughter **Matilda Ledger** headed to the swimming pool. Innocently standing there with no makeup she was remarkably the antithesis of **Marilyn Monroe**. I told her she so deserved the Oscar for her mesmerizing transformation which did not cheer her up knowing the gold was going to Viola or Meryl.

This year there seemed to be more parties than ever. *Vanity Fair* publisher **Edward Menicheschi** staged a staggering six nights of "CAMPAIGN HOLLYWOOD." **Ermenegildo Zenga** and **Colin** and **Livia Firth** hosted an intimate dinner at the Chateau Marmont to benefit Oxfam America, Colin's pet charity. Editor **Graydon Carter** and Edward greeted **Cameron Diaz**, **Kristin Davis**, **Gary** and **Alexandra Oldman** and **Mia Wasikowska.** In addition, Livia spoke about her 'Green Carpet Challenge' which uses eco-friendly fabrics for "wear it once" gowns at awards shows. Get it? Go green on red.

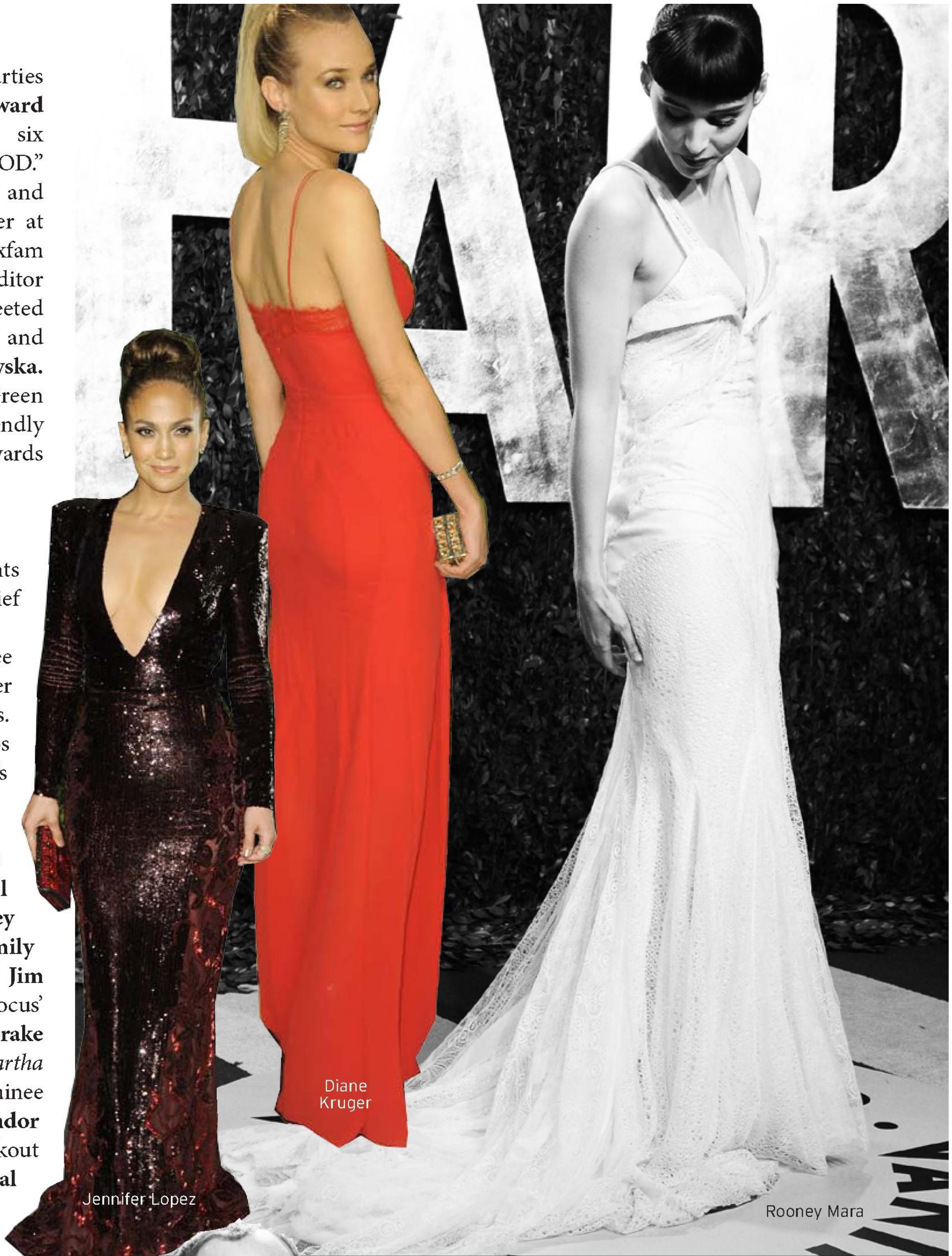
THURSDAY, FEBRUARY 23

Thursday night boasted 15 events causing party panic. Here is a brief rundown of seven.

At *The Hollywood Reporter's* Nominee Night, editor **Janice Min** and publisher **Lynne Segall** greeted the power brokers. With ballots in, competing studios cordially mingled in the Mayor's backyard. **Owen Wilson** slipped in the back door and hung with **Michael Sheen** and producer **Letty Aronson**. Producers **Kathy Kennedy** and **Frank Marshall** chatted with DreamWorks' partner **Stacey Snyder,** producer **Graham King** and **Emily Mortimer**. Fox's **Tom Rothman** and **Jim Gianopulos** compared notes with Focus' **James Shamus**. Young directors **Drake Doremus** (*Like Crazy*), **Sean Durkin** (*Martha Marcy May Marlene*) and Oscar nominee and Spirit Award winner **J.C. Chandor** (*Margin Call*) drank at the bar. Breakout directors **Nick Jarecki** (*Arbitrage*), **Zal Batmanglij** (*Sound of My Voice* and **The East**) and **Jay Duplass** (*Jeff Who Lives at Home*) dreamed about their future nominations. **Aaron Sorkin**, **Piers Morgan** and **Lawrence O'Donnell** handicapped Romney vs. Obama as **Brooklyn Decker** sashayed by.

Urs Fisher's exhibition *Beds & Problem Paintings* featured two bed sculptures at **Larry Gagosian**'s Gallery followed by his private dinner at Mr. Chow's. Art lovers **Vera Wang**, **Russell Simmons**, **Steve Martin**, **Jean Pigozzi** and **John Waters** attended.

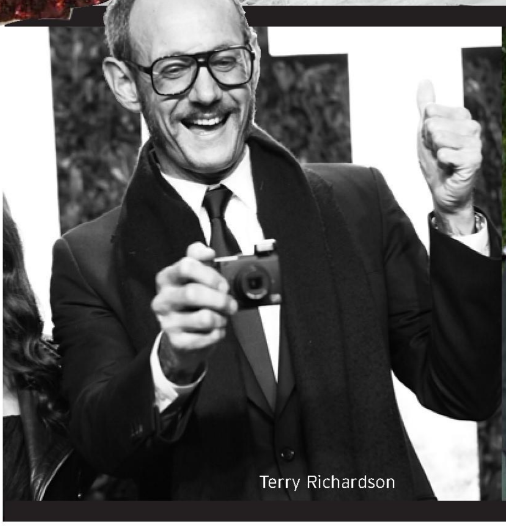
The US-Ireland Alliance honored nominees *Hugo* screenwriter **John Logan,** *Bridesmaids* star **Melissa McCarthy** and Michelle Williams at Bad Robot. Logan also wrote *Rango, Coriolanus*, 007's *Skyfall* and *Jersey Boys* for the big screen.



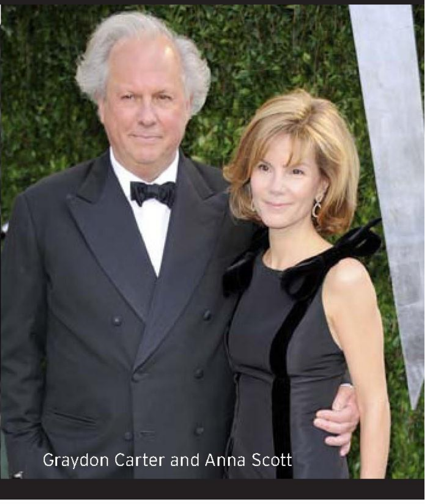
Jennifer Lopez

Diane Kruger

Rooney Mara



Terry Richardson



Graydon Carter and Anna Scott

Michael Douglas and Kirk Douglas

Wendi Deng Murdoch

Fran Leibowitz

Abbie Cornish

Bar Refaeli

Woody is not a member of the Academy because he doesn't feel that films should be in competition. He told me, "A statue does not change your life. You still get a cold. You can't get a date. You still have everyday things to worry about."

Alfre Woodard hosted a down and dirty girls night out in a rented house above Sunset for **Viola Davis** and **Octavia Spencer.**

Universal honcho **Ron Meyer** hosted a civilized private buffet at his Malibu home for Graydon Carter with **Tom Cruise**, **Leonardo DiCaprio**, **Barbra Streisand**, **Tom Hanks** and **Michael Douglas**.

Stunning socialite **Betsey Bloomingdale** gave a seated dinner at her Holmby Hills home for best friends **Nancy Reagan**, **Wendy Stark**, **Bob Colacello**, **Joan Collins** and fashion icon **Lynn Wyatt**.

Tobias Meyer, auctioneer for Sotheby's and art dealer **Mark Fletcher** hosted an open house at their Mulholland Drive home for American avant-garde photographer **Terry Richardson**. This is the only pre-Oscar party where a guest dropped his pants and mooned the red carpet and Terry signed a fans breast. Art collectors **Bill** and **Maria Bell**, **Todd Eberle**, rock singer **Jack Donahu**e and **Francesco Clemente** schmoozed.

FRIDAY, FEBRUARY 24

At the BHH I ran into **David Heyman,** English producer of the *Harry Potter* franchise who was honored at the Publicists Awards lunch at The Beverly Hilton. "Motion Picture Showman of the Year" was the consolation prize for being snubbed by the Academy for visualizing a publishing miracle for children around the world.

"Nobbs" whisked me off to the British Film Reception hosted by **Jeremy Hunt**, UK Secretary of State for Culture and Olympics and the British Consul-General **Dame Barbara Hay**, in her Hancock Park residence. Upon introduction, I blurted out that my friend **Lord Astor** was interested in having L.A. people get to know his son-in-law, **Prime Minister David Cameron**. As an appointed diplomat she was horrified by my indiscretion and turned to greet the next American idiot. I was just making conversation.

Daldry told Sony's **Sir Howard Stringer** and astute film CEO **Michael Lynton, Kenneth Branagh, Janet McTeer** and Gary Oldman that he, as executive producer of the Olympics, was headed back to London to oversee special events, including the opening ceremony, directed by **Danny Boyle**. **Victoria Beckham** made a dramatic, sullen, last-minute appearance looking perfectly skinny in a dress from her eponymous collection.

At the Women in Film cocktail party at Cecconi's, **Gwyneth Paltrow**, **Shailene Woodley**, **Selena Gomez** and **Vanessa Hudgens** networked with Jessica, Octavia and Viola now of social stamina fame.

Blythe Danner kissed me at the door as a military-type looked on. I kept saying to him, "Where have we met?" Nowhere. He was astronaut **Mark E. Kelly** who came with Blythe and is married to former congresswoman **Gabrielle Giffords**, Paltrow's second cousin. Only I could mistake an astronaut for a movie marketing guy.

Felicity Jones

Brooklyn Decker

Miley Cyrus

Vanity Fair fêted Scorsese and The Film Foundation, which has saved 555 films in 22 years. Cocktails were at the restored Bel-Air Hotel. Honorary Jewess **Lorraine Bracco** ran past me yelling that she was late for **Ronald Perelman**'s Shabbat dinner. Three, three-time Oscar winners: composer **Howard Shore**, costume designer **Sandy Powell** and editor **Thelma Schoonmaker** were honored. **Sir Ben Kingsley, Danny Huston**, **Patty Clarkson**, **Irwin Winkler** and **Giorgio Armani**'s niece **Roberta Armani** with **Wanda McDaniels** debated best director: Marty or Michel?

Zoe Saldana

"Nobbs" delivered me to WME's Party at kahuna **Ari Emanuel**'s Brentwood estate, where NFL quarterback and new client **Tim Tebow** was the toast of the party, especially to **Taylor Swift** who made $35.7 million this year. **Michael Douglas** gave me a kiss… doesn't get any better. Longtime lovebirds **Robert Pattinson** and **Kristen Stewart** made a rare appearance, glued to each other's hips. They mingled with co-star Taylor Lautner, **Miley Cyrus,** and her *Hunger Games* beau **Liam Hemsworth**. Meanwhile **Charlize Theron**, **Jack Black**, **Rooney Mara**, **Ben Stiller**, **Barry Sonnenfeld** and **Larry David** talked business with moguls **Les Moonves** and Viacom's **Philippe Dauman**.

Next was UTA Chairman **Jim Berkus**' soiree that police almost shut down because the DJ got carried away impressing **Harrison Ford**, **Channing Tatum**, **Jerry Bruckheimer**, **Tom Freston**, Disney's **Rich Ross**, SNL's **Lorne Michaels** and Oscar show producer **Brian Grazer.**

Sunset Tower Hotel owner **Jeff Klein** and producer **John Goldwyn** hosted a secret dinner for **Anna** and **Graydon Carter** at their Hollywood Hills home with **Tom Ford**, **Mitch Glazer**, **Fran Lebowitz**, **Vito Schnabel**, **Denise Hale,** **Lisa** and **Eric Eisner** and VF's **Punch Hutton,** who is **Tim Hutton**'s sister.

Last stop was CAA **Byran Lourd**'s "Friday Night Party". "Nobbs" was instructed to drop me off at a neighborhood school where a luxury van transported guests to the stone and glass Bel-Air estate situated on a narrow street. I knew that guests **Colin Firth**, **Penelope Cruz**, **Sofia Vergara**, **Salma Hayek,** **Sandra Bullock**, and *Glee*'s **Matthew Morrison** did not arrive by bus. Once inside the playing field leveled out and pound for pound there was more famous flesh per square inch than the Oscars themselves.

I huddled with Meryl on the couch and we talked about her race. She thought Viola. I thought Meryl. She didn't know about Harvey's last-minute "29 year" shout out.

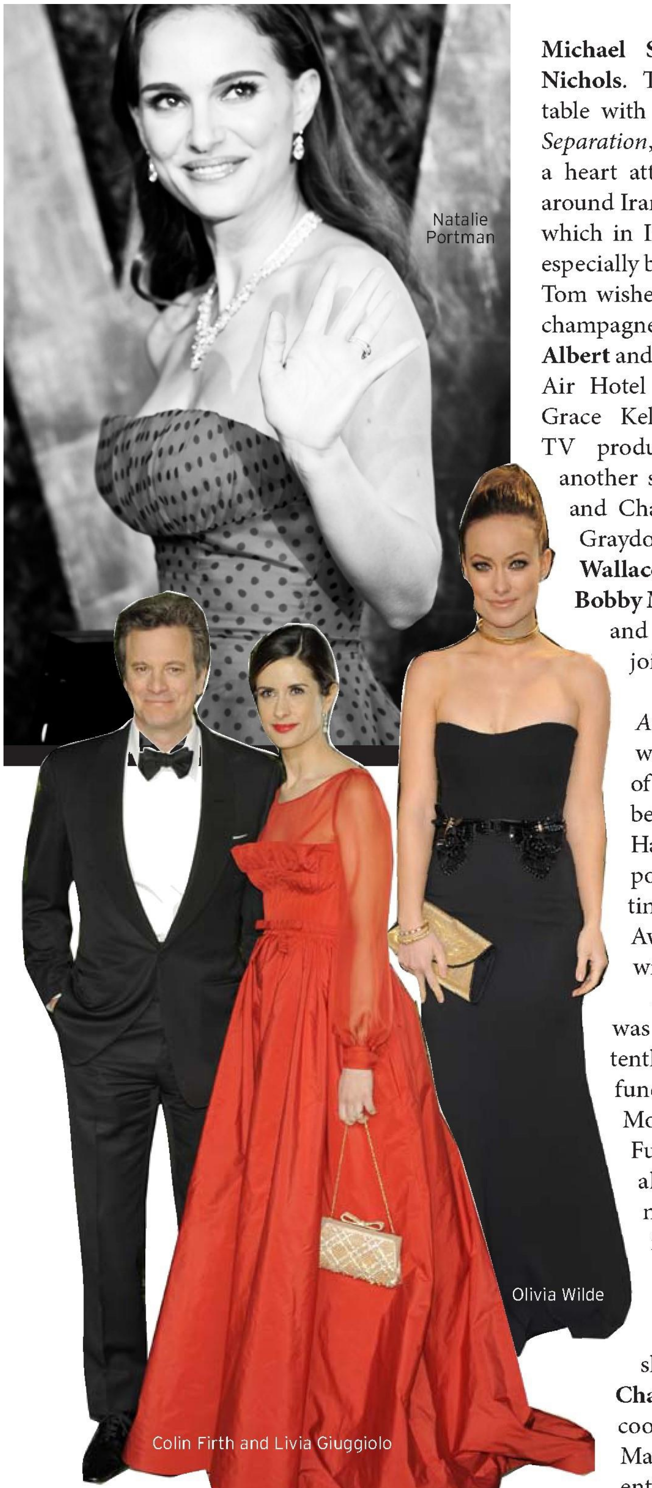
I hugged **Bette Midler** and flirted with **Jim Sturgess** and **Bennett Miller.** Universal's **Donna Langley,** who is overseeing **Tom Hooper**'s production *Les Miserables*, mentioned **Hugh Jackman**'s impeccable manners should insure best behavior from Russell Crowe.

I introduced HBO's *Game Change* director **Jay Roach**, who is an authority on Hitler, to **George Clooney** who is writing a thriller about the Nazis stealing art. Clooney whispered, "The Frenchman is winning."

I thanked **Bryan Lourd**, **Kevin Huvane** and **Richard Lovett**, got on the bus and prayed that I get invited back.

SATURDAY, FEBRUARY 25

I dragged my tired ass to the Academy, as foreign film aficionado **Mark Johnson** was conducting a symposium on the nominated films, which included Sony



Natalie Portman

Olivia Wilde

Colin Firth and Livia Giuggiolo

Classic's *Footnote* from Israel and *A Separation* from Iran. From the other side of the world both governments monitored their directors as the Sony boys kept the peace.

Michael and **Tom Bernard** invited me to the Independent Spirit Awards at the Santa Monica Pier. Michael hosted his *Take Shelter* nominees **Jessica Chastain, Michael Shannon** and director **Jeff Nichols.** Tom held court at the next table with the entire Iranian cast of *A Separation*, which won. Tom almost had a heart attack when I threw my arms around Iranian director **Asghar Farhadi,** which in Iran is unacceptable behavior, especially by a Jewish American Princess. Tom wished I had gone to the crowded champagne brunch in honor of **Prince Albert** and **Princess Charlene** at the Bel-Air Hotel where Montblanc launched Grace Kelly watches. Simultaneously, TV producer **Gary Pudney** hosted another secretive, lunch which Albert and Charlene actually attended with Graydon Carter, **Carolina Herrera, Wallace Annenberg, Bobby Shriver, Bobby Marx, Kathy** and **Ricky Hilton** and **Lynn Wyatt. Wolfgang Puck** joined them for dessert.

"The French," as the *The Artist* gang was nicknamed, had won six Césars, France's version of the Oscars, in Paris the night before. They flew all night and Harvey's chauffeur arranged a police escort from LAX just in time for them to win four Spirit Awards, cementing the Oscar win.

Back at the BHH, Spielberg was the first to arrive at the tenth annual "Night Before" fundraiser in support of the Motion Picture and Television Fund. **Jeffrey Katzenberg** had already secured $200 of a $350 million fundraising goal that included money from him, **Tom Cruise, Steve Bing, Casey Wasserman,** Clooney and Spielberg. Every nominee showed up.

Chanel and **Charles Finch** cooked up their chic soiree at Madeo, where a mariachi band enthusiastically announced everyone's arrival. Bedecked exclusively in Chanel were **Diane Kruger, Elizabeth Olsen, Rose Byrne, Ginnifer Goodwin** and **Rachel Bilson.** Also air kissing were **Rachel Zoe, Rosanna Arquette, Alice Evans** and **Ioan Gruffudd, Zachary Quinto, Julia Ormond** and **Dustin Hoffman.**

My Week With Marilyn's English director **Simon Curtis** insisted we join **Kenneth Branagh** at The Weinstein Company's bash at the Soho House in time to hear **Tony Bennett** sing "Autumn Leaves" to Harvey's surprised 86-year-old mother **Miriam, Madonna** and **Meryl. Jean Dujardin, Bérénice Bejo, Michel Hazanavicius** and producer **Thomas Langmann** staggered around completely jet lagged, too tired to speak English. **Uggie** on the other hand was the absolute star of the evening as his trainer placed him in everyone's arms for photos. *W.E*'s **Andrea Riseborough** and **Abbie Cornish** sat next to baseball cap-clad **Leonardo DiCaprio** as his ex, **Bar Refaeli,** kept her distance across the room. **Zoe Saldana** walked in holding hands with **Bradley Cooper. Scarlett Johansson** introduced me to her boyfriend **Nate Naylor. Katy Perry, Felicity Jones** and **Malin Akerman** circled a refreshed **Gerard Butler.** The two daughters of New York slain hero cop **Peter Figoski,** Corrine, 14 and Caroline, 16 who had also been Harvey's guests at the Super Bowl stood in the middle of this circus and just fainted.

SUNDAY, FEBRUARY 26

I met interior designer **Nicky Haslam** in the Polo Lounge and found **Nancy Reagan** brunching with **Bob Colacello** and **Carolina Herrera.** Nancy knows me as "The DVD Lady". I promised to send her *The Iron Lady.* I didn't have the heart to tell her that her late husband, who was **Margaret Thatcher**'s best friend, was barely mentioned in the film.

In a Marchesa gown, Dennis Basso fur and Iradj Moini necklace, I collected **Simon Curtis** and headed to the Hollywood & Highland Center. Simon, who has never been to the Oscars before, miraculously scored a front row seat between Michelle Williams and Clooney. We pulled up to screams of hysteria at the mother of all red carpets.

Meryl's publicist **Leslee Dart** whispered to me. "She is dressed like an Oscar. What do we do if she loses?" **Sacha Baron Cohen** hilariously guilted the Academy into letting him wear his costume from *The Dictator* and, after soiling **Ryan Seacrest,** he went directly to dinner at *Vanity Fair.* **Gwyneth Paltrow** won the style award in Tom Ford's white column and **Angelina Jolie** so successfully invented a new one legged pose in a thigh high slit gown on the red carpet she repeated it for 39.3 million

I introduced HBO's Game Change director Jay Roach, who is an authority on Hitler, to George Clooney who is writing a thriller about the Nazis stealing art. Clooney whispered, "The Frenchman is winning."

people onstage.

Tickets were so tight that I gave my plus one to **Penelope Ann Miller** because she promoted *The Artist* every day for four months. Seated next to us were co-stars **James Cromwell** and **Missi Pyle**.

Billy Crystal made us laugh, Cirque de Soleil made us gasp and most of the wins were expected.

Colin Firth crowned Meryl her third win on a record 17 nominations and Harvey beamed.

The Academy was so confident *The Artist* would win, they invited Uggie, who waited in the wings and ran out as **Tom Cruise** announced the film.

Once the show was off the air, I followed my seatmates to the stage loaded with their programs, wraps and handbags and led the French to the press rooms.

The Governor's Ball was the next stop where the winners got their Oscars engraved. Everyone paid respects to the Academy's **Tom Sherak** and **Dawn Hudson**. Stars headed to the Sunset Tower to Graydon's glittering Vanity Fair Oscar bash and their militarized security with micro chipped cards. If your name was not on the list and you were carrying an Oscar, you could walk in. **Billy** and **Janice Crystal** were mobbed with well wishers. **Elton John** made $5 million dollars at his 20th AmFar event which also auctioned off two tickets to Vanity Fair's party for $230,000.**Tom Cruise** and **Katie Holmes** were thrilled to talk to **George Lucas**. Gwyneth snuggled with Coldplay's **Chris Martin** and **Jennifer Lopez** brought boy toy **Casper Smart.** Every major actor previously mentioned is standing in the room. Spotted in a sea of celebrity were **Tina Fey**, **Glenn Close**, **Olivia Wilde** and **Jason Sudeikis**, **Terry George, Jane Fonda**, **Demian Bichir**, **Claire Danes,** Meryl's daughters **Mamie Gummer** with husband **Ben Walker**, **Grace Gummer**, **Salma Hayek**, **Brit Marling, Natalie Portman**, **Sofia Coppola**, **Peter Brant**, **Stephanie Seymour**, **Ryan Kavanaugh**, **Ingrid Sischy**, **Sandy Brant** and **Wendi Murdoch** with **Bingbing Li**.

George Clooney threw his own exclusive after party at Craig's in West Hollywood for close friends **Bryan Lourd**, **Grant Heslov**, **Stan Rosenfield**, **Brad** and **Angelina**, **Emily Blunt** and **John Krasinski**, **Cindy Crawford**, **Jimmy Kimmel**, **Ryan Seacrest** and best adapted screenplay winner **Alexander Payne**.

George and Brad cancelled each other out with two great performances. Clooney immediately returned to his best role as humanitarian, flew to the Sudan, met with Obama and was dramatically arrested at a protest.

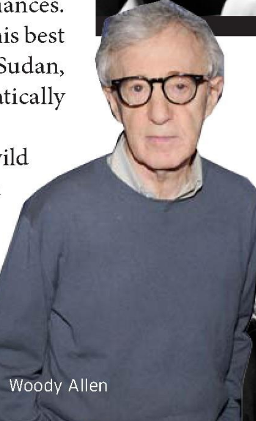
"The French" threw a wild celebration at the Chateau Marmont, poured champagne down their throats and threw each other in the pool at 3:00 a.m. Nobody spoke English.

By 4:00 a.m. Harvey rounded up his artists for a live broadcast on *The Today Show* from The Four Seasons lobby. In disheveled black tie, five Oscar winners, all total unknowns a year ago, made Oscar history with the first silent film to win since 1927.

It wasn't God, but 5,800 Academy voters who said they won. This had to be the most exciting night of their lives. The glory and memories live forever. The next day they went back to reality . As Woody says, "A statue does not really change your life. You still get a cold."

Uggie was invited to the White House correspondent's dinner as a guest of the Washington Times in April. He hopes to meet President Obama. ✦

Jean Dujardin



Woody Allen

Elizabeth Olsen

Claire Danes

Georgina Chapman and Harvey Weinstein

From: Boris Nikolic [redacted]
Sent: 3/23/2012 7:10:58 PM
To: Jeffrey Epstein (jeevacation@gmail.com) [jeevacation@gmail.com]
Subject: FW: World Bank Nominee
Attachments: image005.jpg; image006.jpg

Importance: High

The Atlantic Home
Friday, March 23, 2012

- Politics

Massoud Hayoun - Massoud Hayoun writes for and produces *The Atlantic*'s International channel.

How Jim Yong Kim, Obama's World Bank Pick, Changed Global Health Aid

By Massoud Hayoun
Mar 23 2012, 2:04 PM ET

The international public health work that made Kim, now the president's nominee for World Bank head, such a respected figure.

President Obama introduces Jim Yong Kim as his nominee to be the next president of the World Bank / Reuters

President Obama announced today that he will nominate Dartmouth College President Jim Yong Kim to head the World Bank. Although Kim is a physician by training, officials have observed that Kim's role as a key player in global health and development, notably with his role in the organization Partners in Health (PIH), makes him a key candidate to change the attitudes of developing-world nations.

The following excerpts from Tracy Kidder's *Mountains Beyond Mountains*, a biography of anthropologist and physician Paul Farmer, detail Kim's bold efforts to combat international HIV and tuberculosis epidemics with PIH:

Some months after the official founding of PIH, [co-founder] Paul Farmer expanded the group, adding a fellow Harvard anthropology and medical student, a Korean American named Jim Yong Kim... Farmer offered what for Jim Kim was a convincing vision of the new organization. The reality was less impressive -- a charity with a board of advisers and no hired staff...

They talked about issues such as political correctness, which Jim Kim defined as follows: "It's a very well-crafted tool to distract us. A very self-centered activity. Clean up your own vocabulary so you can show everybody you have the social capital of having been in circles where these things are talked about on a regular basis." (What was an example of political correctness? Some academic types would say to Jim and Paul, "Why do you call your patients poor people? They don't call themselves poor people." Jim would reply: "Okay, how about soon-dead people?")

They talked about the insignificance of "cultural barriers" when it came to the Haitian peasant's acceptance of modern Western medicine: "There's nothing like a cure for a disease to change people's cultural values"...

By now Peru was taxing PIH's resources severely. On average, the drugs to treat just one patient cost between fifteen and twenty thousand dollars. And the number of patients kept growing. Already there were about fifty Carabayllanos in treatment. Their average age was twenty-nine. They were students, unemployed youths, housewives, street vendors, bus drivers, health workers. The actual numbers seemed small, but those fifty MDR [a form of tuberculosis that does not respond to standard treatment] cases represented about 10 percent of all active cases of TB in the slum, about ten times more than might have been expected. No telling how many others they had been infecting as they'd traveled around Lima, coughing. No telling either how many people in other parts of the city already had MDR, but [there were] reports of hundreds in other neighborhoods. In Carabayllo itself, the Socios workers found entire families sick and dying with what turned out to be genetically related strains of the disease--a phenomenon common enough that the health workers gave it a name, familias tebeceanas, tuberculosis families.

Kim's organization confronted Peru's MDR-form tuberculosis epidemic with what some have called unorthodox practices -- borrowing and cajoling its way into medicine for its patients.

Howard Hiatt, a friend of Jim Yong Kim's and a former dean at the Harvard School of Public Health, said he was concerned about how PIH was getting medicine to combat the epidemic:

"Sure enough. Paul and Jim would stop at the [Harvard-affiliated] Brigham pharmacy before they left for Peru and fill their briefcases with drugs. They had sweet-talked various people into letting them walk away with the drugs." [Hiatt] was amused, all in all. "That's their Robin Hood attitude." In fact, they'd only borrowed the drugs...

Then one day the president of the Brigham stopped Hiatt in a corridor. "Your friends Farmer and Kim are in trouble with me. They owe this hospital ninety-two thousand dollars." Hiatt looked into the matter. "Sure enough. Paul and Jim would stop at the Brigham pharmacy before they left for Peru and fill their briefcases with drugs. They had sweet-talked various people into letting them walk away with the drugs." He was amused, all in all. "That's their Robin Hood attitude."

To many seasoned managers of public health projects, what Farmer and Kim were doing would have looked quite reckless--like a stunt, as some would later insinuate. They didn't have a guaranteed supply of drugs, only the determination to obtain the drugs and the charm to get away with borrowing. They were borrowing their laboratory services, too, from Massachusetts. They lacked proper institutional support. The weight of expert opinion stood against them. Their organization was small and it had other projects, in Haiti and Boston and elsewhere, and Peru put a strain on everyone. Jim had to travel to Carabayllo at least once a month. Farmer had to go there slightly more often.

Kim's audacious 'Robin Hood attitude' won him and PIH acclaim for their role in changing global health and development.

In June 2002 ... the WHO adopted new prescriptions for dealing with MDR-TB, virtually the same as PIH had used in Carabayllo. For Jim Kim this marked the end of a long campaign. "The world changed yesterday," he wrote from Geneva to all of PIH. The prices of second-line antibiotics continued to decline, and the drugs now flowed fairly smoothly through the Green Light Committee to, among other places, Peru, where about 1,000 chronic patients were either cured or in treatment. About 250 were receiving the drugs in Tomsk, and, largely because of the efforts of WHO, the Russian Ministry of Health had finally agreed to the terms of the World Bank's TB loan--150 million dollars to begin to fight the epidemic throughout the country.

The twin pandemics of AIDS and tuberculosis raged on, of course, magnifying each other, in Africa and Asia, Eastern Europe and Latin America. Mathematical models predicted widening global catastrophe--100 million HIV infections in the world by the year 2010. Some prominent voices, some in the U.S. government, still argued that AIDS could not be treated in desperately impoverished places. But this view seemed to be fading. The prices of antiretrovirals were falling, even more dramatically than the prices of second-line TB drugs.

This was thanks to a growing worldwide campaign for treating AIDS wherever it occurred. Jim Kim had often said that the world's response to AIDS and TB would define the moral standing of his generation. In 2003, a new director general took over at WHO, and he asked Jim to serve as his senior adviser. Meanwhile, the example of Zanmi Lasante [PIH's Haiti-based project] was growing, and Cange had become a favorite destination for global health policy makers and American politicians.

AFFIDAVIT OF BRADLEY JAMES EDWARDS

1. I am an attorney in good standing with the Florida Bar and admitted to practice in the Southern District of Florida. I am a partner in the law firm of Farmer Jaffe Weissing Edwards Fistos and Lehrman.

2. I am the lead attorney currently representing "Jane Doe" in the case of Jane Doe v. Jeffrey Epstein, case number 08-80893 in federal Court in the Southern District of Florida. I am the lead attorney representing Jane Doe, whose civil complaint alleges that Epstein sexually molested her numerous occasions when she was a minor.

3. Defendant Epstein has entered into a "non-prosecution agreement" (NPA) with the federal government for sex crimes against minors. Under that agreement, the federal government has agreed not to file criminal charges against Epstein for sex crimes committed against approximately thirty girls, including Jane Doe. In exchange, Epstein agreed to plead guilty to state law criminal charges involving solicitation of prostitution and procuring a minor for prostitution. The victim of the criminal charges to which he has pled was not Jane Doe.

4. Under the NPA, Epstein has agreed not to contest civil liability of any of his approximately thirty victims – provided that the victim agrees to limit themselves to the damages provided by 18 U.S.C. § 2255 (currently set at $150,000). Jane Doe has not agreed to limit herself to pursuing only $150,000 in damages. Therefore, the terms of the NPA purport to prevent Jane Doe from using the NPA to prove liability.

5. Epstein has filed an answer to Jane Doe's complaint, in which he has invoked his Fifth Amendment right to silence with respect to the allegations that he molested her as a child. Epstein has further argued that this Fifth Amendment invocation is the functional equivalent of, and must be treated as, a specific denial of the allegations.

6. Defendant Epstein's deposition has been taken on several occasions, in this and other related cases, and he has not provided any substantive discovery whatsoever. Instead, he invoked his 5th amendment privilege against self-incrimination when asked questions about his abuse of Jane Doe or other girls.

7. Defendant Epstein has also been served with Interrogatories and requests for production; all requests have been met with 5th amendment assertions and Epstein has not given Jane Doe any substantive testimony related her allegations.

8. Jane Doe's complaint contains a punitive damages claim, and Mr. Epstein has also elected to invoke the 5th Amendment on all questions that would relate to punitive damages issues, such as his intent when committing the crimes, his lack of remorse and his intent to recidivate.

9. Epstein has taken Jane Doe's deposition. During that deposition he has asked numerous questions of Jane Doe that suggest that she is fabricating her allegation of abuse by Epstein.

10. In addition to deposing Mr. Epstein, other attorneys and I have taken the depositions of his various co-conspirators (as labeled by the federal government in the NPA), including [redacted], [redacted] and [redacted]. Each of those individuals was employed by Epstein to bring him underage girls for him to molest and to ensure that he was protected from detection by law enforcement, and thus those individuals could likely provide general testimony that would assist Plaintiff in proving liability and damages, including punitive damages. However, none of these individuals were

present during acts of sexual abuse by Epstein. In any event, ALL of those individuals have also invoked their 5th amendment rights against self-incrimination, and thus have left Plaintiff with no information about what Epstein or other conspirators inside his house were doing during the sexual abuse of Jane Doe and other minors girls. This creates a serious issue for Jane Doe in proving her sexual molestation claim against Epstein. By its nature, sexual molestation takes place in private, with only the abuser and the victim typically available to testify. In this case, Epstein's abuse of Jane Doe took place in private, with only Epstein and Jane Doe present during the abuse. Jane Doe has no other reasonable avenues of discovery to provide direct proof of claim of sexual abuse by Epstein.

11. Additionally, Mr. Epstein has recently filed a lawsuit against me personally that has no merit whatsoever, a fact known to Mr. Epstein and his attorneys. He filed the lawsuit against Brad Edwards, Scott Rothstein, and [redacted](another Epstein victim of his molestation). That lawsuit implies that L.M.'s civil case against him (currently pending in Florida state court) is fabricated and that [redacted] and I have conspired to commit fraud against him (presumably that she made up the case against him, implying that he does not know [redacted]. While the present subpoena before the Court has been filed by Jane Doe, the Court should be aware that attorneys representing [redacted] may also file a subpoena for the George Rush tape shortly.

12. Despite Mr. Epstein and all of his co-conspirators, asserting a 5th amendment privilege against self-incrimination, George Rush of the New York Daily news did contact me to inform me that Mr. Epstein spoke personally with him about issues related to the various charges of sex abuse against him.

13. Paraphrasing from memory of my conversation with Mr. Rush, Mr. Epstein told him that he may have come "too close to the line" but that he should not have been punished as severely as he was and that his conduct was at most worthy of a $100 fine. This is a statement that shows two things of great importance to Jane Doe's pending civil action. First, it is in effect an admission by Epstein of his liability to Jane Doe for sexually abusing her. Jane Doe does not have any other admission of Epstein of his sexual abuse of her and Epstein has filed an answer to Jane Doe's complaint that has the functional effect of denying abuse of her. Jane Doe has diligently pursued all possible ways of obtaining an admission from Epstein of his molestation of Jane Doe without success. Second, the statement to Mr. Rush is a clear demonstration that Epstein lacks remorse for committing felony child molestation against Jane Doe. This will be a central issue in the punitive damages case against Epstein at trial. Here again, Jane Doe has diligently pursued all possible ways of obtaining a statement from Epstein about his lack of remorse for abusing Jane Doe without success. There are no other reasonable means of obtaining a statement from Epstein on these subjects.

14. Mr. Rush also told me that Mr. Epstein spoke specifically about one of my clients, [redacted] and he made derogatory remarks about her.

15. Additionally, Mr. Rush said that Epstein spoke directly about another civil case that was filed against him (Jane Doe 102 v. Epstein); that case alleges that Epstein repeatedly sexually abused a 15 year old girl, forced her to have sex with his friends and flew her on his private plane nationally and internationally for the purposes of sexually molesting and abusing her. Epstein flippantly told George Rush that that case was dismissed, in a way to indicate that the allegations are ridiculous and untrue.

16. Mr. Rush indicated that he taped the conversation between him and Mr. Epstein.

17. Mr. Rush also spoke at length to Michael Fisten, an investigator with my firm that was assisting with the investigation of the case. Mr. Fisten reported to me shortly after the conversation with Mr. Rush that he had such a conversation.

18. While research by other plaintiffs' attorneys and myself has uncovered other persons that were acquaintances of Mr. Epstein, specifically Donald Trump, Alan Dershowitz, Bill Clinton, Tommy Mottola, and David Copperfield, we have no information that any of those people (other than Mr. Dershowitz) have spoken to Mr. Epstein about Jane Doe or any of the other specific victims of Mr. Epstein's molestation. Mr. Dershowitz is acting as an attorney for Mr. Epstein, and therefore it is presumably unlikely to question him about any admissions that Epstein may have made regarding Jane Doe or other minors girls. Additionally, we have no information that any of those individuals or any other individuals have any taped statements of Epstein's own voice relating to these matters. George Rush's taped conversation with Mr. Epstein is the only known one in existence, making it very unique and it contains information not otherwise obtainable through other means or sources. Indeed, without the Rush tape conversation, the jury that handles the case will not hear any words from Epstein himself about his abuse of Jane Doe and other young girls. I have been informed by Epstein's attorney that Epstein intends to invoke his Fifth Amendment rights rather than answer any substantive questions about the abuse of Jane Doe and other girls at trial.

19. The Rush interview is, in any event, unique and not otherwise obtainable from other witnesses because it can be used to prove perjury (a federal crime) on the part of Epstein. Epstein lied about not knowing George Rush. See deposition of Jeffrey Epstein, taken in [redacted] v. Jeffrey Epstein, case 50-2008-CA-028051, page 154, line 4 through 155 line 9, wherein Jeffrey Epstein clearly impresses that he does not recognize George Rush from the New York Daily News, despite the fact that he gave a personal interview that we all now know to have been tape recorded. It is therefore evidence of a criminal event. If we receive the tape, we intend to alert the appropriate law enforcement authorities, both federal and state, so that they can pursue any appropriate criminal investigation perjury charges.

20. The tape is also crucial for [redacted] to dismiss the frivolous complaint filed by Jeffrey Epstein against her, as he clearly acknowledges knowing [redacted] contrary to claims he makes in his complaint against her and also contradictory to other statements he has made in depositions related to knowing [redacted] In that regard, this tape provides evidence of other false statements Epstein has made under oath.

21. During a telephone call with George Rush, he provided me more than a description of the tape, and in fact described the general tenor of the entire interview, so that nothing in the interview can be fairly regarded as confidential at this point.

22. As George Rush admitted in his affidavit, he played the tape for *at least* two other persons who also confirmed Epstein's arrogance as he speaks about his actions with minors.

23. The people for whom George Rush played the tape or told in detail of the information on the tape were not "sources" in the tradition sense of the word – all individuals were simply chatting with Mr. Rush about Mr. Epstein and his propensity to molest children. For example, when I discussed the tape with Mr. Rush, I was not a "source" in the traditional sense of that term. At no point did Mr. Rush tell me that I was a "source" for his reporting.

24. Because Epstein and all other co-conspirators have invoked the 5th amendment as to all relevant questions, this tape is the *only* way that Jane Doe can put Epstein's own perceptions of what he has done before the jury and the only way that Jane Doe can put Epstein's admissions and statement s before the jury. As even a quick perusal of the more than 500 entries on the docket sheet for Jane Doe's (consolidated) case will confirm (see Case no. 9:08-80119 (S.D. Fla.) (case number for consolidated cases on discovery), Jane Doe and other plaintiffs have made exhaustive attempts to obtain information from Epstein about his abuse. These attempts have included repeated requests for admission, requests for production, interrogatories, and depositions – all the means that are listed in the Federal Rules of Civil Procedure for obtaining discovery. These means have all been exhausted without success. Neither

Jane Doe nor any of the other plaintiffs have been able to obtain even a single word of information from Epstein about his abuse of minor girls.

25. I made a good faith, albeit unsuccessful, effort to resolve this matter with Anne B. Carroll, representing the Daily News in order to avoid any court intervention. I explained that we needed this tape for several reasons, including those cited by her in her pleading. The tape is detrimental to Epstein's personal complaint against [redacted] and me; the tape is evidence of perjury committed by Epstein; the tape is the Best Evidence of his lack of remorse for his actions and will be presented in the punitive damages phase of the civil trials against him; and, perhaps most important, the tape is the only way that the jury considering Jane Doe's case will be able to hear Epstein's voice and own statements about his abuse of Jane Doe and other minor girls. Without the tape, the jury will not have the opportunity to hear Epstein give any substantive information about Jane Doe's complaint. Indeed, they will not have the opportunity to even hear Epstein's voice utter any substantive words other than (in essence) "I take the Fifth." As part of our discussion, Ms. Carroll told me that it was a "stupid move" for Mr. Rush to play the tape or disclose the tape to other people as he likely waived any privilege and that, as a result of disclosing the tape, he was at risk of losing his job. I responded that it did not seem fair that Mr. Rush lose his job or be punished in any way, but that I had an absolute duty to represent my client and that I would be failing in that duty if I did not pursue this critical piece of evidence.

I declare under penalty of perjury that the foregoing is true and correct.

FURTHER AFFIANT SAYETH NAUGHT.

Dated this 23rd day of April, 2010.

Brad Edwards, Esq.

The foregoing instrument was acknowledged before me this 23rd day of April, 2010 by BRAD EDWARDS, who is personally known to me.

NOTARY PUBLIC

Print Name:____________________

My Commission Expires:

MARIA W. KELLJCHIAN
MY COMMISSION # DD 813393
EXPIRES: August 11, 2012
Bonded Thru Notary Public Underwriters

EXHIBIT N

AFFIDAVIT OF BRADLEY JAMES EDWARDS

1. I am an attorney in good standing with the Florida Bar and admitted to practice in the Southern District of Florida. I am currently a partner in the law firm of Farmer, Jaffe, Weissing, Edwards, Fistos & Lehrman, P.L.

2. In 2008, I was a sole practitioner running a personal injury law firm in Hollywood, FL. While a sole practitioner I was retained by three clients, [redacted], [redacted], and Jane Doe to pursue civil litigation against Jeffrey Epstein for sexually abusing them while they were minor girls. I agreed to represent these girls, along with attorney Jay Howell (an attorney in Jacksonville, Florida with Jay Howell & Associates) and Professor Paul Cassell (a law professor at the University of Utah College Of Law). I filed state court actions on behalf of L.M. and E.W. and a federal court action on behalf of Jane Doe. All of the cases were filed in the summer of 2008.

3. My clients received correspondence from the U.S. Department of Justice regarding their rights as victims of Epstein's federal sex offenses. (True and accurate copies of the letters are attached to Statement of Undisputed Facts as Exhibit "M")

4. In mid June 2008, I contacted Assistant United States Attorney Marie Villafaña to inform her that I represented Jane Doe #1[redacted].) and, later, Jane Doe #2([redacted]). I asked to meet to provide information regarding Epstein. AUSA Villafaña did not advise me that a plea agreement had already been negotiated with Epstein's attorneys that would block federal prosecution. AUSA Villafaña did indicate that federal investigators had concrete evidence and information that Epstein had sexually molested at least 40 underage minor females, including[redacted]., Jane Doe and [redacted]

5. I also requested from the U.S. Attorney's Office the information and evidence that they had collected regarding Epstein's sexual abuse of his clients. However, the U.S. Attorney's Office declined to provide any such information to me. The U.S. Attorney's Office also declined to provide any such information to the other attorneys who represented victims of Epstein's sexual assaults.

6. I was informed that on Friday, June 27, 2008, at approximately 4:15 p.m., AUSA Villafaña received a copy of Epstein's proposed state plea agreement and learned that the plea was scheduled for 8:30 a.m., Monday, June 30, 2008. She called me to provide notice to my clients regarding the hearing. She did not tell me that the guilty pleas in state court would bring an end to the possibility of federal prosecution pursuant to the plea agreement. My clients did not learn and understand this fact until July 11, 2008, when the agreement was described during a hearing held before Judge Marra on the Crime Victims' Rights Act action that I had filed.

7. In the summer of 2008 I filed complaints against Jeffrey Epstein on behalf of[redacted], E.W., and Jane Doe.

8. In the Spring of 2009 (approximately April), I joined the law firm of Rothstein, Rosenfeldt and Adler, P.A. ("RRA"). I brought my existing clients with me when I joined RRA, including [redacted] [redacted]., and Jane Doe. When I joined the firm, I was not aware that Scott Rothstein was running a Ponzi scheme at RRA. Had I known such a Ponzi scheme was in place, I would never have joined RRA.

9. I am now aware that it has been alleged that Scott Rothstein made fraudulent presentations to investors about the lawsuits that I had filed on behalf of my clients against Epstein and that it has been alleged that these lawsuits were used to fraudulently lure investors into Rothstein's Ponzi scheme. I never met a single investor, had no part in any such presentations and had no knowledge any such fraud was occurring. If these allegations are true, I had no knowledge that any such fraudulent presentations were occurring and no knowledge of any such improper use of the case files.

10. Epstein's Complaint against me alleges that Rothstein made false statements about cases filed against Epstein, i.e., that RRA had 50 anonymous females who had filed suit against Epstein; that Rothstein sold an interest in personal injury lawsuits, reached agreements to share attorneys fees with non-lawyers, paid clients "up front" money; and that he used the judicial process to further his Ponzi scheme. If Rothstein did any of these things, I had no knowledge of his actions. Because I maintained close contact with my clients, [redacted], [redacted] and Jane Doe, and Scott Rothstein never met any of them, I know for certain that none of my clients were paid "up front" money by anyone.

11. Epstein alleges that I attempted to take the depositions of his "high profile friends and acquaintances" for no legitimate litigation purpose. This is untrue, as all of my actions in representing [redacted]., [redacted], and Jane Doe were aimed at providing them effective representation in their civil suits. With regard to Epstein's friends, through documents and information obtained in discovery and other means of investigation, I learned that Epstein was sexually molesting minor girls on a daily basis and had been for many years. I also learned the unsurprising fact that he was molesting the girls in the privacy of his mansion in West Palm Beach, meaning that locating witnesses to corroborate their testimony would be difficult to find. I also learned, from the course of the litigation, that Epstein and his lawyers were constantly attacking the credibility of the girls, that Epstein's employees were all represented by lawyers who apparently were paid for (directly or indirectly) by Epstein, that co-conspirators whose representation was also apparently paid for by Epstein were all taking the Fifth (like Epstein) rather than provide information in discovery. For example, I was given reason to believe that [redacted], Larry Visoski, Larry Harrison, David Rogers, Louella Rabuyo, [redacted] [redacted], Ghislaine Maxwell, Mark Epstein, and Janusz Banasiak all had lawyers paid for by Epstein. Because Epstein and the co-conspirators in his child molestation criminal enterprise blocked normal discovery avenues, I needed to search for other ordinary approaches to strengthen the cases of my clients. Consistent with my training and experience, these other ordinary approaches included finding other witnesses who could corroborate allegations of sexual abuse of my clients or other girls. Some of these witnesses were friends of Epstein. Given his social status, it also turned out that some of his friends were high-profile individuals.

12. In light of information I received suggesting that British socialite Ghislaine Maxwell, former girlfriend and long-time friend of Epstein's, was involved in managing Epstein's affairs and companies I had her served for deposition for August 17, 2009. (Deposition Notice attached to Statement of Undisputed Facts as Exhibit BB). Maxwell was represented by Brett Jaffe of the New York firm of Cohen and Gresser, and I understood that her attorney was paid for (directly or indirectly) by Epstein. She was reluctant to give her deposition, and I tried to work with her attorney to take her deposition on terms that would be acceptable to both sides. Her attorney and I negotiated a confidentiality agreement, under which Maxwell agreed to drop any objections to the deposition. Maxwell, however, still avoided the deposition. On June 29, 2010, one day before I was to fly to NY to take Maxwell's deposition, her attorney informed me that Maxwell's mother was deathly ill and Maxwell was consequently flying to England with no intention of returning and certainly would not return to the United States before the conclusion of Jane Doe's trial period (August 6, 2010). Despite that assertion, I later learned that Ghislaine Maxwell was in fact in the country on approximately July 31, 2010, as she attended the wedding of Chelsea Clinton (former President Clinton's daughter) and was captured in a photograph taken for US Weekly magazine.

13. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Donald Trump's deposition in the civil suits against him. Trump was properly noticed because: (a) after review of the message pads confiscated from Epstein's home, the legal and investigative team assisting my clients learned that Trump called Epstein's West Palm Beach mansion on several occasions during the time period most relevant to my clients' complaints; (b) Trump was quoted in a *Vanity Fair* article about Epstein as saying **"I've known Jeff for fifteen years. Terrific guy." "He's a lot of fun to be with. It is even said that he likes beautiful women as much as I do, and many of them are on the younger side. No doubt about it – Jeffrey enjoys his social life."** Jeffrey Epstein: International Moneyman of Mystery; **He's pals with a passel of Nobel Prize–winning scientists, CEOs like Leslie Wexner of the Limited, socialite Ghislaine Maxwell, even Donald Trump. But it wasn't until he flew Bill Clinton, Kevin Spacey, and Chris Tucker to Africa on his private Boeing 727 that the world began to wonder who he is. By Landon Thomas Jr.**; (c) I learned through a source that Trump banned Epstein from his Maralago Club in West Palm Beach because Epstein sexually assaulted an underage girl at the club; (d) Jane Doe No. 102's complaint alleged that Jane Doe 102 was initially approached at Trump's Maralago by Ghislaine Maxwell and recruited to be Maxwell and Epstein's underage sex slave; (e) Mark Epstein (Jeffrey Epstein's brother) testified that Trump flew on Jeffrey Epstein's plane with him (the same plane that Jane Doe 102 alleged was used to have sex with underage girls) deposition of Mark Epstein, September 21, 2009 at 48-50; (f) Trump visited Epstein at his home in Palm Beach – the same home where Epstein abused minor girls daily; (g) Epstein's phone directory from his computer contains 14 phone numbers for Donald Trump, including emergency numbers, car numbers, and numbers to Trump's security guard and houseman. Based on this information, I believed that

Trump might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

14. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Alan Dershowitz's deposition in the civil suits against him. Dershowitz was properly noticed because: (a) Dershowitz has been friends with Epstein for many years; (b) in one news article Dershowitz comments that, "I'm on my 20th book… The only person outside of my immediate family that I send drafts to is Jeffrey" The Talented Mr. Epstein, By Vicky Ward on January, 2005 in Published Work, Vanity Fair; (c) Epstein's housekeeper Alfredo Rodriguez testified that Dershowitz stayed at Epstein's house during the years most relevant to my clients; (d) Rodriguez testified that Dershowitz was at Epstein's house at times when underage females where there being molested by Epstein (see Alfredo Rodriguez deposition at 278-280, 385, 426-427); (e) Dershowitz was reportedly involved in persuading the Palm Beach State Attorney's office not to file felony criminal charges against Epstein because the underage females lacked credibility and thus could not be believed that they were at Epstein's house, despite him being an eyewitness that the underage girls were actually there; (f) Jane Doe No. 102 stated generally that Epstein forced her to be sexually exploited by not only Epstein but also Epstein's "adult male peers, including royalty, politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories that Dershowitz and acquaintances of Dershowitz fall into; (g) during the years 2002-2005 Alan Dershowitz was on Epstein's plane on several occasions according to the flight logs produced by Epstein's pilot and information (described above) suggested that sexual assaults may have taken place on the plane; (h) Epstein donated Harvard $30 Million dollars one year, and Harvard was one of the only institutions that did not return Epstein's donation after he was charged with sex offenses against children. Based on this information, I believed that Dershowitz might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

15. Epstein alleges that there was something improper in the fact that I notified him that I intended to take Bill Clinton's deposition. Clinton was properly noticed because: (a) it was well known that Clinton was friends with Ghislaine Maxwell, and several witnesses had provided information that Maxwell helped to run Epstein's companies, kept images of naked underage children on her computer, helped to recruit underage children for Epstein, engaged in lesbian sex with underage females that she procured for Epstein, and photographed underage females in sexually explicit poses and kept child pornography on her computer; (b) newpaper articles stated that Clinton had an affair with Ghislaine Maxwell, who was thought to be second in charge of Epstein's child molestation ring. The Cleveland Leader newspaper, April 10, 2009; (c) it was national news when Clinton traveled with Epstein (and Maxwell) aboard Epstein's private plane to Africa and the news articles classified Clinton as Epstein's friend; (d) the flight logs for the relevant years 2002 - 2005 showed Clinton traveling on Epstein's plane on more than 10 occasions and his assistant, Doug Band, traveled on many more occasions; (e) Jane Doe No. 102 stated generally that she was required by Epstein to be sexually

exploited by not only Epstein but also Epstein's "adult male peers, including royalty, politicians, academicians, businessmen, and/or other professional and personal acquaintances" – categories Clinton and acquaintances of Clinton fall into; (f) flight logs showed that Clinton took many flights with Epstein, Ghislaine Maxwell, [redacted] and [redacted], -- all employees and/or co-conspirators of Epstein's that were closely connected to Epstein's child exploitation and sexual abuse; (g) Clinton frequently flew with Epstein aboard his plane, then suddenly stopped – raising the suspicion that the friendship abruptly ended, perhaps because of events related to Epstein's sexual abuse of children; (h) Epstein's personal phone directory from his computer contains e-mail addresses for Clinton along with 21 phone numbers for him, including those for his assistant (Doug Band), his schedulers, and what appear to be Clinton's personal numbers. Based on this information, I believed that Clinton might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

16. Epstein alleges that Tommy Mottola was improperly noticed with a deposition. I did not notice Mattola for deposition. He was noticed for deposition by a law firm representing another one of Epstein's victims – not by me.

17. Epstein alleges that there was something improper in the fact that I notified him that I intended to take the illusionist David Copperfield's deposition. Copperfield was properly noticed because: (a) Epstein's housekeeper Alfredo Rodriguez testified that David Copperfield was a guest on several occasions at Epstein's house; (b) according to the message pads confiscated from Epstein's house, Copperfield called Epstein quite frequently and left messages that indicated they socialized together; (c) Copperfield himself has had similar allegations made against him by women claiming he sexually abused them; (d) one of Epstein's sexual assault victims also alleged that Copperfield had touched her in an improper sexual way while she was at Epstein's house. Based on this information, I believed that Copperfield might have relevant information to provide in the cases against Jeffrey Epstein and accordingly provided notice of a possible deposition.

18. Epstein alleges that there was something improper in the fact that I identified Bill Richardson as a possible witness against him in the civil cases. Richardson was properly identified as a possible witness because Epstein's personal pilot testified to Richardson joining Epstein at Epstein's New Mexico Ranch. See deposition of Larry Morrison, October 6, 2009, at 167-169. There was information indicating that Epstein had young girls at his ranch which, given the circumstances of the case, raised the reasonable inference he was sexually abusing these girls since he had regularly and frequently abused girls in West Palm Beach and elsewhere. Richardson had also returned campaign donations that were given to him by Epstein, indicating that he believed that there was something about Epstein that he did not want to be associated with. Richardson was not called to testify nor was he ever subpoenaed to testify.

19. Epstein alleges that discovery of plane and pilot logs was improper during discovery in the civil cases against him. Discovery of these subjects was clearly proper and

necessary because: (a) Jane Doe filed a federal RICO claim against Epstein that was an active claim through much of the litigation. The RICO claim alleged that Epstein ran an expansive criminal enterprise that involved and depended upon his plane travel. Although Judge Marra dismissed the RICO claim at some point in the federal litigation, the legal team representing my clients intended to pursue an appeal of that dismissal. Moreover, all of the subjects mentioned in the RICO claim remained relevant to other aspects of Jane Doe's claims against Epstein, including in particular her claim for punitive damages; (b) Jane Doe also filed and was proceeding to trial on a federal claim under 18 U.S.C. § 2255. Section 2255 is a federal statute which (unlike other state statutes) guaranteed a minimum level of recovery for Jane Doe. Proceeding under the statute, however, required a "federal nexus" to the sexual assaults. Jane Doe had two grounds on which to argue that such a nexus existed to her abuse by Epstein: first, his use of the telephone to arrange for girls to be abused; and, second, his travel on planes in interstate commerce. During the course of the litigation, I anticipated that Epstein would argue that Jane Doe's proof of the federal nexus was inadequate. These fears were realized when Epstein filed a summary judgment motion raising this argument. In respo-nse, the other attorneys and I representing Jane Doe used the flight log evidence to respond to Epstein's summary judgment motion, explaining that the flight logs demonstrated that Epstein had traveled in interstate commerce for the purpose of facilitating his sexual assaults. Because Epstein chose to settle the case before trial, Judge Marra did not rule on the summary judgment motion. (c) Jane Doe No. 102's complaint outlined Epstein's daily sexual exploitation and abuse of underage minors as young as 12 years old and alleged that he used his plane to transport underage females to be sexually abused by him and his friends. The flight logs accordingly might have information about either additional girls who were victims of Epstein's abuse or friends of Epstein who may have witnessed or even participated in the abuse. Based on this information, I believed that the flight logs and related information was relevant information to prove the cases against Jeffrey Epstein and accordingly I pursued them in discovery.

20. In approximately November 2009, the existence of Scott Rothstein's Ponzi scheme became public knowledge. It was at that time that I, along with many other reputable attorneys at RRA, first became aware of Rothstein criminal scheme. At that time, I left RRA with several other RRA attorneys to form the law firm of Farmer Jaffe Weissing Edwards Fistos and Lehrman ("Farmer Jaffe"). I was thus with RRA for less than one year.

21. In July 2010, along with other attorneys at Farmer Jaffe and Professor Cassell, I reached favorable settlement terms for my three clients [redacted]., [redacted]., and Jane Doe in their lawsuits against Epstein.

22. On July 20, 2010, I received a letter from the U.S. Attorney's Office for the Southern District of Florida – the office responsible for prosecuting Rothstein's Ponzi scheme. The letter indicated that law enforcement agencies had determined that I was "a victim (or potential victim)" of Scott Rothstein's federal crimes. The letter informed me of my rights as a victim of Rothstein's federal crimes and promised to keep me informed about

subsequent developments in his prosecution. A copy of this letter is attached to this Affidavit. (A copy of the letter is attached to Statement of Undisputed Facts as Exhibit UU)

23. Jeffrey Epstein also filed a complaint with the Florida Bar against me. His complaint alleged that I had been involved in Rothstein's scheme and had thereby violated various rules of professional responsibility. The Florida Bar investigated and dismissed the complaint.

24. I have reviewed the Statement of Undisputed Facts filed contemporaneously with this Affidavit. Each of the assertions concerning what I learned, what I did, and the good faith beliefs formed by me in the course of my prosecutions of claims against Jeffrey Epstein as contained in the Statement of Undisputed Facts is true, and the foundations set out as support for my beliefs are true and correct to the best of my knowledge.

25. All actions taken by me in the course of my prosecution of claims against Jeffrey Epstein were based upon a good faith belief that they were reasonable, necessary, and ethically proper to fulfill my obligation to zealously represent the interests of my clients.

I declare under penalty of perjury that the foregoing is true and correct.

Dated: 9/21, 2010

Bradley J. Edwards, Esq.

EXHIBIT A

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF FLORIDA

CASE NO.: 08-CIV-80893-MARRA/JOHNSON

JANE DOE,

Plaintiff,

vs.

JEFFREY EPSTEIN,

Defendant.
____________________________________/

DEFENDANT EPSTEIN'S MOTION FOR SETTLEMENT CONFERENCE, OR IN THE ALTERNATIVE, MOTION TO DIRECT PARTIES' BACK TO MEDIATION

Defendant, JEFFREY EPSTEIN, by and through his undersigned attorneys, pursuant to the Federal Rules of Civil Procedure and the Local Rules for the Southern District of Florida, moves this Court for an order requiring the parties to attend a Settlement Conference before Magistrate Judge Linnea R. Johnson, or in the alternative, for an Order directing the parties to reconvene at a second mediation on or before July 1, 2010, and as grounds set forth would state:

1. The above-styled matter is currently scheduled on the Court's trial docket beginning July 19, 2010. (D.E. #119, Order Re-Setting Trial Date and Pretrial Deadlines). The Court's Mandatory Pretrial Stipulation and Motions in Limine deadlines are set for July 1, 2010. In this regard, if the parties could reach an agreement at a settlement conference or a mediation before these pre-trial deadlines, it would result in substantial conservation of judicial resources and preparation time.

2. The parties attended mediation on April 5, 2010, at Matrix Mediation, LLC, with Rodney Romano serving as mediator, but were unable to reach an agreement. (See D.E. #139).